PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : A01N 43/22, A61K 31/71 C07H 17/08, C12N 1/21 C12P 19/62	A1	(11) International Publication Number: WO 93/13663
		(43) International Publication Date: 22 July 1993 (22.07.93)

(21) International Application Number: PCT/US92/00427

(22) International Filing Date: 17 January 1992 (17.01.92)

(71) Applicant: ABBOTT LABORATORIES [US/US]; CHAD D-377/AP6D-2, One Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(72) Inventors: KATZ, Leonard ; 844 North Avenue, Waukegan, IL 60085 (US). DONADIO, Stefano ; 110 Brookhill Road, Libertyville, IL 60048 (US). MCALPINE, James, B. ; 211 West Rockland Road, Libertyville, IL 60048 (US).

(74) Agents: GORMAN, Edward, Hoover, Jr. et al.; Abbott Laboratories, CHAD D-377/AP6D-2, One Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: METHOD OF DIRECTING BIOSYNTHESIS OF SPECIFIC POLYKETIDES

(57) Abstract

A method to produce novel polyketide structures by designing and introducing specified changes in the DNA governing the synthesis of the polyketide is disclosed. The biosynthesis of specific polyketide analogs is accomplished by genetic manipulation of a polyketide-producing microorganism by isolating a polyketide biosynthetic gene-containing DNA sequence, identifying enzymatic activities associated within the DNA sequence, introducing one or more specified changes into the DNA sequence which codes for one of the enzymatic activities which results in an altered DNA sequence, introducing the altered DNA sequence into the polyketide-producing microorganism to replace the original sequence, growing a culture of the altered microorganism under conditions suitable for the formation of the specific polyketide analog, and isolating the specific polyketide analog from the culture. The method is most useful when the segment of the chromosome modified is involved in an enzymatic activity associated with polyketide biosynthesis, particularly for manipulating polyketide synthase genes from *Saccharapolyspora* or *Streptomyces*.

METHOD OF DIRECTING BIOSYNTHESIS OF SPECIFIC POLYKETIDES

Field of the Invention.

5   The present invention relates to a method for directing the biosynthesis of specific polyketide analogs by genetic manipulation. In particuar, polyketide biosynthetic genes are manipulated to produce precise, novel polyketides of predicted structure.

10 Background of the Invention

Polyketides are a large class of natural products that includes many important antibiotics and immunosuppressants such as erythromycins, tetracyclines, and rapamycins. Their synthesis proceeds by an ordered condensation of acyl esters to generate carbon chains of varying length and

15 substitution pattern that are later converted to mature polyketides. This process has long been recognized as resembling fatty acid biosynthesis, but with important differences. Unlike a fatty acid synthase, a typical polyketide synthase is programmed to make many choices during carbon chain assembly: For example, the choice of "starter" and "extender" units,

20 which are often selected from acetate, propionate or butyrate residues in a defined sequence. The choice of using a full cycle of reduction-dehydration-reduction after some condensation steps, omitting it completely, or using one of two incomplete cycles (reduction alone or reduction followed by dehydration), which determines the pattern of keto

25 or hydroxyl groups and the degree of saturation at different points in the chain is additionally programed. Finally the choice of stereochemistry for the substituents at many of the carbon atoms is programmed by the polyketide synthase.

Because of the commercial significance of *Streptomyces*, a great

30 amount of effort has been expended in the study of *Streptomyces* genetics. Consequently much is known about *Streptomyces* and several cloning vectors exist for transformations of the organism.

Although many polyketides have been identified, there remains the need to obtain novel polyketide structures with enhanced properties.

35 Current methods of obtaining such molecules include screening of natural isolates and chemical modification of existing polyketides, both of which are costly and time consuming. Current screening methods are based on gross properties of the molecule, i.e. antibacterial, antifungal

activity, etc., and both *a priori* knowledge of the structure of the molecules obtained or predetermination of enhanced properties are virtually impossible. Chemical modification of preexisting structures has been successfully employed, but it still suffers from practical limitations to the

5 type of compounds obtainable, largely connected to the poor yield of multistep syntheses and available chemistry to effect modifications. The following modifications are extremely difficult or inefficient at the present time: change of the stereochemistry of the side chains in the completed polyketide; change of the length of the polyketide by removal or addition

10 of carbon units from the interior of the acyl chain; and dehydroxylation at unique positions in the acyl chain. Accordingly, there exists the need to obtain molecules wherein such changes can be specified and performed and would represent an improvement in the technology to produce altered polyketide molecules with predicted structure.

15

Summary of the Invention

The present invention provides a method to produce novel structures from designing and introducing specified changes in the DNA governing the synthesis of the polyketide. According to the method of the

20 present invention, the biosynthesis of specific polyketide analogs is accomplished by genetic manipulation of a polyketide-producing microorganism comprising the steps of:

(1) isolating a polyketide biosynthetic gene-containing DNA

25 sequence;

(2) identifying enzymatic activities associated within said DNA sequence;

(3) introducing one or more specified changes into said DNA sequence which codes for one of said enzymatic activities which results in

30 an altered DNA sequence;

(4) introducing said altered DNA sequence into the polyketide-producing microorganism to replace the original sequence;

(5) growing a culture of the altered microorganism under conditions suitable for the formation of the specific polyketide analog; and

35 (6) isolating said specific polyketide analog from the culture.

The present method is most useful when the segment of the chromosome modified is involved in an enzymatic activity associated with polyketide biosynthesis. The present invention is especially useful

3

in manipulating polyketide biosynthetic genes from *Streptomyces*, an organism which provides over one-half of the clinically useful antibiotics.

Brief Description of the Drawings

5   FIG. 1 illustrates the organization of gene encoding polyketide synthase and designated *eryA* as follows: (a) Map coordinates of the DNA; (b) DOTPLOT of the output of COMPARE (window = 50, stringency = 32) program (Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin, Biotechnology Center, 1710 University
10   Avenue, Madison, Wisconsin 53705 of *eryA* segment (x-axis) vs. subsegment of *eryA* comprises between 23 - 27.5 sequence coordinates (y-axis) [see Fig. 2]; (c) Open reading frame organization of *eryA* and enzymatic activities encoded. PT = propionyltransferase; ACP = acyl carrier protein; KS=β-ketoacyl ACP synthase; RmT = (2R) methylmalonyl
15   CoA transferase; KR =β-ketoredu:ase; SmT = (2S) methylmalonyl CoA transferase; DH = dehydratase; EK = enoylreductase; TE = thioesterase; and (d) Schematic diagram showing the extent of each of the six modules in *eryA*.

          FIG. 2. illustrates the nucleotide sequence of *eryA* with
20   corresponding translation of the three open reading frames. Standard one letter codes for the amino acids appear beneath their respective nucleic acid codons. The standard one letter codes for the amino acid sequences are as follows:

25   A   -alanine
                        R   -arginine
                        N   -asparagine
                        D   -aspartic acid
                        C   -cysteine
30   Q   -glutamine
                        E   -glutamic acid
                        G   -glycine
                        H   -histidine
                        I   -isoleucine
35   L   -leucine
                        K   -lysine
                        M   -methionine (start)
                        F   -phenylalanine

P	-proline
S	-serine
T	-threonine
W	-tryptophan
Y	-tyrosine
V	-valine

FIG. 3. is a schematic representation of Type I, Type II and Type III changes in *eryA* and structures of corresponding novel polyketides produced. Δ69 (Type I) and Δ33 (Type II) represent in-frame deletions of the base pairs in the DNA segments corresponding to the KR of module 2 and the β—ketoacyl ACP synthase of module 2, respectively. Insertion of a complete copy of module 4 within module 1 is also shown. Production of 11-epifluoro-15-norerythromycin in strain that carries Δ33 occurs when substrate analog (2S,3S,4S,5S)2,4-dimethyl-3-fluoro-5-hydroxyhexanoic acid-ethyl thioester is fed.

FIG. 4 illustrates the restriction site coordinates of cosmid pR1 5' to the sequence of *eryA* (Fig 2).

Detailed Description of the Invention

For the purposes of the present invention as disclosed and claimed herein, the following terms are defined.

The term "polyketide" as used refers to a large and diverse class of natural products, including antibiotics, pigments, and immunosuppressants. Antibiotics include, but are not limited to anthracyclines, tetracyclines, polyethers, ansamycins, macrolides of different types (polyenes and avermectins as well as classical macrolides such as erythromycins).

The term "polyketide-producing microorganism" as used herein includes any *Actinomycetales* which can produced a polyketide. Examples of Actinomycetes that produce polyketides include but are not limited to *Micromonospora rosaria*, *Micromonospora megalomicea*, *Sacharapolyspora erythraea*, *Streptomyces antibioticus*, *Streptomyces albireticuli*, *Streptomyces ambofasciens*, *Streptomyces avermitilis*, *Streptomyces fradiae*, *Streptomyces hygroscopicus*, *Streptomyces tsukubaensis*, *Streptomyces griseus*, *Streptomyces mycarofasciens*, *Streptomyces platensis*, *Streptomyces venezuelae*, *Streptomyces*

violaceoniger, and various *Actinomadura, Dactylosporangium* and *Nocardia* strains that produce polyether type of polyketides.

The term "polyketide synthase" as used herein refers to the complex of enzymatic activities responsible for the biosynthesis of

5   polyketides which include but are not limited to β-ketoreductase, dehydratase, acyl carrier protein, enoylreductase, β-ketoacyl ACP synthase, and acyltransferase.

The term "extender" as used herein refers to a coenzyme A thioester of a dicarboxylate which is incorporated into a polyketide by a

10   polyketide synthase.

The term "starter" as used herein refers to a coenzyme A thioester of a carboxylic acid which is used by the polyketide synthase as the first building block of the polyketide.

The term "*eryA*" as used herein refers to the genes involved in the

15   formation of the polyketide moiety of erythromycin.

The term "condensation" as used herein refers to the addition of an extender unit out to the nascent polyketide chain and requires the action of β-ketoacyl ACP synthase, acyltransferase, and acyl carrier protein.

The term "β-carbonyl processing" as used herein refers to changes

20   effecting the carbonyl group of the growing polyketide via β-ketoreductase, dehydratase, and enoylreductase.

The term "module" as used herein refers to the genetic element encoding one condensation step, as defined above, and one β-carbonyl processing step, as defined herein.

25   The term "Type I change" as used herein refers to changes in DNA sequence which will result in the production of polyketide rings of length identical to that of 6-deoxyerythronolide A, but with altered functional groups at specific ring positions.

The term "Type II change" as used herein refers to alterations

30   which will result in the production of macrolide rings only when fed exogenously with substrate analogs, e.g. thioesters of appropriate acyl compounds of various length. Thus Type II mutants are erythromycin non-producing (Ery⁻) mutants. The structure of the resulting macrolides will depend on the substrate employed.

35   The term "Type III change" as used herein refers to alterations which will result in the biosynthesis of macrolide rings of length reduced (deletion) or increased (insertion) by two carbon units, or macrolide rings altered in specific portions of the chain (replacement).

6

In its broadest sense, the present invention entails a general procedure for producing novel polyketide structures *in vivo* by selectively altering the genetic information of the organism that naturally produces a related polyketide. A set of examples described herein are a

5   series of novel polyketides that make use of the genetic information for the biosynthesis of the polyketide portion of the macrolide antibiotic erythromycin. The organization of the segment of the *Saccharapolyspora erythraea* chromosome, designated *eryA*, and the corresponding polypeptides which it encodes that determine the biosynthesis of the

10   polyketide segment of erythromycin, are shown in FIG. 1. It is seen that *eryA* is organized in modules, as shown, and that each module takes care of one condensation step, through the action of the β-ketoacyl ACP synthase specified within, wherein an extender unit, methylmalonyl CoA, is added first to the starter unit, propionyl CoA, and then to the

15   successively growing acyl chain. The precise succession of elongation steps is dictated by the genetic order of the six modules: module 1 determines the first condensation; module 2, the second; module 3, the third, and so on until the sixth condensation step has occurred. Furthermore, the processing of the growing chain after each condensation

20   is also determined by the information within each module. Thus β-ketoreduction of the β-carbonyl takes place after each step except for step 3, as determined by the presence of a functional β-ketoreductase in all modules except module 3, whereas dehydration and enoylreduction only take place after the fourth extender unit is added to the growing acyl chain,

25   as determined by the presence of dehydratase and enoylreductase in module 4. Furthermore, the choice of the correct enantiomer (2R or 2S) of methylmalonyl-CoA as the extender unit employed at each condensation is specified by the acyltransferase function determined by each module (FIG. 1C).

30   In the present invention, novel polyketide molecules of desired structure are produced by the introduction of specific genetic alterations of the *eryA* sequence into the *Sac. erythraea* chromosome. The complete nucleotide sequence of the *eryA* segment of the *Sac. erythraea* chromosome and the sequence of the corresponding polypeptides are

35   shown in FIG. 2. Three types of alterations to the *eryA* DNA sequence are described: (i) those inactivating a single function in a module which does not arrest acyl chain growth (β-ketoreductase, dehydratase, or enoylreductase); (ii) those inactivating a single functi n in a module

which does arrest chain growth (β–ketoacyl ACP synthase, acyltransferase or acyl carrier protein); and (iii) those affecting an entire module (deletion, insertion, or replacement). The novel polyketides produced by strains carrying these types of mutations can be classified accordingly. Type I

5 changes will result in the production of polyketide rings of length identical to that of 6-deoxyerythronolide A, but with altered functional groups at specific ring positions. Strains carrying type II alterations will result in the production of macrolide rings only when fed exogenously with substrate analogs, e.g.thioesters of appropriate acyl compounds of

10 various length. Thus Type II mutants are erythromycin non-producing (Ery⁻) mutants. The structure of the resulting macrolides will depend on the substrate employed. Type III changes will result in the biosynthesis of macrolide rings of length reduced (deletion) or increased (insertion) by two carbon units, or macrolide rings altered in specific portions of the

15 chain (replacement). A schematic representation of some examples of Type I, Type II and Type III alterations in *eryA* and the corresponding novel polyketides produced in hosts that carry such alterations is shown in FIG. 3.

In the examples described herein, specific mutations in the *eryA*

20 region of the *Sac. erythraea* chromosome are introduced by a simple two-step approach: 1) introduction of a specified change in a cloned DNA segment; 2) exchange of the wild type allele with the mutated one. Step 1 requires standard recombinant DNA manipulations employing *E. coli* as the host. Step 2 requires one or more plasmids out of the several *E. coli-*

25 *Sac. erythraea* shuttle vectors available and a simple screening procedure for the presence of the colony carrying the altered gene. Two methods are used to introduce the altered allele into the chromosome to replace the wild type allele. The first employs gene replacement, described in Examples 7, 11, 15, 19 and 24, wherein the gene to be altered, along with

30 adjacent upstream and downstream DNA, is mutated and cloned into a *Sac. erythraea* non-replicating vector. The plasmid carrying the altered allele is then introduced into the host strain by transformation of protoplasts employing selection for a plasmid marker. Since the plasmid does not replicate, regenerated cells that carry the marker have undergone

35 a single homologous recombination between one of the two segments flanking the mutation on the plasmid and its homologous counterpart in the chromosome. Some of the colonies that have subsequently lost the marker will have undergone a second recombination between the other

plasmid borne adjacent DNA segment and its homologous chromosomal counterpart resulting in the retention of the mutation in the chromosome, replacing the normal allele with the mutant one. The second method to introduce an altered allele into the chromosome

5   employs gene conversion, described in Examples 37 and 43. In this method, an Ery⁻ *Sac. erythraea* strain carrying a deletion of a specified region of the *eryA* segment of the chromosome is used as a host. Into a *Sac. erythraea* multicopy plasmid that carries a selectable marker is cloned the wild type counterpart (segment 1) of the *eryA* segment mutant in the

10   host. Subsequently, the desired homologous or heterologous DNA segment to be introduced (segment 2) is cloned within the portion of segment 1 which is deleted in the mutant strain. The resulting plasmid is then introduced into the host employing selection for the marker. Among the transformants will be a population that have integrated

15   segments 1 and 2 from the plasmid by the process of gene conversion which can be verified by examination of the DNA among the colonies that have recovered the ability to produce erythromycin.

Two examples each of Types I, II and III alterations to the *eryA* DNA sequence and the resultant novel polyketides produced are described in

20   the examples described herein. Examples 1 through 8, 9 through 12 and 13 through 16 describe the construction and effect of three Type I mutants. Examples 17 through 22 and 23 through 27 describe the construction of two Type II mutants and the effects of feeding two different synthetic substrates to the mutant strains. Examples 28 through 38 and 39 through

25   44 outline the steps in constructing Type III changes and their respective effects on the structure of the novel polyketides produced. In Examples 1 through 7 a plasmid that contains a substantial deletion of the segment of the gene corresponding to the b-ketoreductase of module 5 is created, the altered gene is inserted into the *Sac. erythraea* chromosome to replace the

30   wild type allele and the new strain carrying the altered gene is identified and isolated. In Example 8, the new strain is fermented and the novel polyketide 5-oxo-5,6-dideoxy-3α-mycarosyl erythronolide B that results from the introduction of the mutant allele is isolated. In Examples 9 through 11, a mutation is introduced into the β−ketoreductase of module

35   2 and the mutated allele is then used to replace the wild type allele in the chromosome. In Example 12, the strain carrying the altered allele is fermented and the novel compound 11-oxo-11-deoxyerythromycin A is isolated. Similarly, in Examples 13 through 16 a mutation is introduced

9

into the dehydratase of module 4 and the mutated allele is then used to replace the wild type allele in the chromosome. The strain carrying this altered allele is then fermented and the novel products 7-hydroxyerythromycin A and 6-deoxy-7-hydroxyerythromycin A are

5   isolated. In Examples 17 through 21, a mutation is made in the DNA corresponding to the β-ketoacyl-ACP synthase of module 1 and introduced into the chromosome to replace the wild type allele. This mutation has the effect of arresting the synthesis of the polyketide chain and results in the Ery⁻ phenotype. The synthetic substrate (2S,3R,4S,5S)3,5-dihydroxy-2,4-

10   dimethylhexanoic acid-ethyl ester is then made and fed to the mutant resulting in the production of the novel compound (14S,15S)14(1-hydroxyethyl)erythromycin. Similarly, in Examples 22 through 24, a mutation is created in the β-ketoacyl-ACP synthase of module 2 and introduced into the chromosome to replace the wild type allele. In

15   Example 25 and 26, the synthetic substrate (2S,3S,4S,5S)2,4-dimethyl-3-fluoro-5-hydroxyhexanoic acid-ethyl thioester is made and fed to the module 2 β-ketoacyl-ACP synthase mutant and the resulting novel compound 11-epifluoro-15-norerythromycin is isolated. In Examples 27 through 38, a copy of the DNA sequence corresponding to module 4 is

20   introduced into the deleted segment of the β-ketoacyl-ACP synthase of module 1 resulting in the production of the novel compound 14(1-propyl)erythromycin. In Examples 40 through 44, a copy of the DNA sequence corresponding to module 5 is introduced into the deleted segment of the β-ketoacyl ACP synthase of module 1 resulting in the

25   production of the novel compound 14[1(1-hydroxypropyl)]erythromycin.

GENERAL METHODS

30   <u>Materials, Plasmids and Bacterial Strains</u>

        Restriction endonucleases, T4 DNA ligase, nick-translation kit, competent E. coli DH5α cells , X-gal, IPTG, and plasmids pUC19 and pUC12 are purchased from Bethesda Research Laboratories (BRL), Gaithersburg, MD. [α-³²P]dCTP and Hybond N are from Amersham Corp., Chicago, IL.

35   Seakem LE agarose and Seaplaque low gelling temperature agarose are from FMC Bioproducts, Rockland, ME. E. coli K12 strains carrying the E. coli-Sac. shuttle plasmids pWHM3 or pWHM4 (Vara et al.,J. Bacteriol., 171: 5872 (1989)) or the cosmids pS1 (Tuan et al.,Gene, 90: 21 (1990)) and

Sac. erythraea strain NRRL2338 have been deposited in the culture collection of the Agricultural Research Laboratories, Peoria, IL and are available under the accession numbers NRRL XXXX, respectively. Staphylococcus aureus Th^R (thiostrepton resistant) is obtained by plating

5   10^8 cells of S. aureus on agar medium containing 10 mg/ml thiostrepton and picking a survivor after 48 hr growth at 37°C. Thiostrepton is obtained from Squibb-Bristol Myers, New Brunswick, NJ. All other chemical and reagents are from standard commercial sources unless specified otherwise.

10

DNA Manipulations

        Standard conditions (Maniatis et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982) are employed for restriction endonuclease digestion, agarose

15   gel-electrophoresis, nick translation of DNA to make ^{32}P-labeled probes, DNA ligation, and transformation of E. coli employing selection for ampicillin resistance (Ap^R) on LB agar plates. Plasmid DNA is isolated from minipreps of E. coli transformants by the boiling method (Maniatis et al., 1982, supra). DNA fragments are recovered from low melting

20   agarose gels using the method of Langridge et al., 1980. Total DNA from Sac. erythraea strains is prepared according to described procedures (Hopwood et al., Genetic Manipulation of Streptomyces, A Laboratory Manual, John Innes Foundation, Norwich, U.K., 1985). DNA is transferred from agarose gels onto Hybond N following the

25   manufacturer's instructions. Hybridizations are performed in sealed bags containing 10-20 ml of [1xNET (20xNET = 3 M NaCl, 0.3 M TrisHCl, 20 mM Na$_2$EDTA, pH 8.0), 5XDenhardt's solution (Maniatis et al., 1982, supra), 0.2 mg/ml denatured calf thymus DNA, 0.2% SDS, and 0.5-2x10^7 cpm of the nick-translated probe] for 16-20 hr at 65 °C. Filters are washed three times

30   in 1xNET/0.1% SDS for 20 min each at room temperature, and once in 0.05xNET/0.1% SDS for 20 min at 70 °C. Filters are reused as described (Donadio et al., 1990).

Amplification of DNA fragments

35   Synthetic deoxyoligonucleotides are synthesized on an ABI Model 380A synthesizer (Applied Biosystems, Foster City, CA) following the manufacturer's recommendations. Amplification of DNA fragments is performed by the polymerase chain reaction (PCR) employing a Coy

thermocycler. Reactions contain 100 pmol c ach primer, 1 μg of template
DNA (cosmid pS1 carrying the *eryA* segmen from <u>Sac</u>. <u>erythraea</u> strain
NRRL 2338), and 2.5 units of <u>Thermus aquaticus</u> DNA polymerase in a 100
ml volume of PCR buffer [50 mM KCl, 10 mM TrisHCl (pH 8.0) 2 mM

5      $MgCl_2$, 0.01% gelatin) containing 200 mM of the 4 dNTPs. The above
reagents are from Perkin Elmer Cetus, Norwalk, CT. The reaction mixture
is overlaid with a drop of paraffin oil and subjected to 30-50 cycles. Each
cycle consists of one 94 °C, one 55 °C and one 72 °C period, each of the
duration of 3 min. The progress of the amplification is monitored by

10   agarose gel-electrophoresis. The PCR primers described in the examples
below are derived from the nucleotide sequence of *eryA* of FIG. 2.

Gene replacement and gene conversion

       Protoplasts of *Sac. erythraea* strains are prepared and transformed

15   with miniprep DNA isolated from <u>E</u>. <u>coli</u> according to published
procedures (Yamamoto et al., 1986). Integrative transformants, in the case
of pWHM3 derivatives, are selected after one round of non-selective
growth of the primary Th^R transformants as described by Weber et. al,
<u>Gene</u>, 68: 173 (1988). Loss of the Th^R phenotype is monitored by plating

20   serial dilutions of a Th^R integrant on non-selective medium, followed by
replica-plating on thiostrepton-containing medium. Th^S (thiostrepton-
sensitive) colonies arise at a frequency of 10^{-2} (Donadio et al., 1990). The
retention of the mutant allele is established by Southern hybridization of a
few Th^S colonies.

25   - A few hundred Th^R colonies obtained by transformation of an *eryA*
strain with pWHM4 derivatives are screened for antibiotic production by
the agar-plug assay employing *Staphylococcus aureus* as Th^R organism as
described (Tuan et al.,<u>Gene</u>, 90: 21 (1990)). The frequency of gene
conversion between a 5 kb segment of homologous sequence and a strain

30   carrying a small deletion is >25% (Tuan et al.,<u>Gene</u>, 90: 21 (1990)). Colonies
found to produce antibiotic activity are inoculated in SGGP (Yamamoto <u>et</u>
<u>al</u>., 1986), protoplasts are prepared, and the regenerated protoplasts are
scored for loss of the plasmid by replica-plating on non-selective medium.
Th^S colonies are then rechecked for antibiotic production, and six

35   producers are analyzed on Southern blots.

<u>Fermentation</u>

Sac. erythraea cells are inoculated into 100 ml SCM medium (1.5% soluble starch, 2.0% Soytone [Difco], 0.15% Yeast Extract [Difco], 0.01% $CaCl_2$) and allowed to grow at 32°C for 3 to 6 days. The entire culture is
5   then inoculated into 10 liters of fresh SCM medium. The fermenter is operated for a period of 7 days at 32°C maintaining constant aeration and pH at 7.0. After fermentation is complete, the cells are removed by centrifugation at 4°C and the fermentation beer is kept in the cold until further use.
10   The present invention will now be illustrated, but is not intended to be limited, by the following examples:

<div align="center">

Example 1
<u>Construction of plasmid pABX9</u>
</div>

15

The 9.6 kb <u>BamHI</u>-<u>XhoI</u> segment comprised between sequence coordinates 21.96 and 31.52 was isolated from cosmid pS1 and ligated to <u>SalI</u>-digested pUC19 DNA. The resulting mixture contained the desired plasmid pABX9.

20

<div align="center">

Example 2
<u>Construction of E. coli K12 DH5α/pABX9</u>
</div>

Approximately 10 ng of plasmid pABX9, prepared as described in
25   Example 1, were transformed into E. coli K12 DH5α and a few of the resulting white Ap^R colonies that appeared on the LB-agar plates containing X-gal and ampicillin were analyzed for their plasmid content. One colony was found to carry pABX9, as verified by the observation of fragments of 3.93, 3.39, 2.01, 1.56, 0.87, and 0.48 kb in size upon agarose gel
30   electrophoresis after <u>SmaI</u> digestion of the plasmid.

<div align="center">

Example 3
<u>Construction of plasmid pABX9DN</u>
</div>

35   Plasmid pABX9, isolated from E. coli K12 DH5α/pABX9, was digested with <u>NcoI</u> and then treated with T4 DNA ligase. The resulting mixture contained the desired plasmid pABX9DN.

Example 4
Construction of E. coli K12 DH5a/pABX9DN

Approximately 10 ng of plasmid pABX9DN, prepared as described
in Example 3, were transformed into E. coli K12 DH5α and a few of the
resulting white Ap^R colonies that appeared on the LB-agar plates
containing X-gal and ampicillin were analyzed for their plasmid content.
Colonies carrying pABX9DN exhibited a single NcoI fragment of 11.5 kb
visible by agarose gel electrophoresis, confirming that the 813 bp NcoI -
NcoI fragment from pABX9 has been deleted in pABX9DN.

Example 5
Construction of plasmid pABX95DN

Plasmid pABX95DN was digested with EcoRI and HindIII and
ligated to pWHM3 digested with the same two enzymes. The resulting
mixture contained the desired plasmid pABX95DN.

Example 6
Construction of E. coli K12 DH5α/pABX95DN

Approximately 10 ng of plasmid pABX95DN, prepared as described
in Example 5, were transformed into E. coli K12 DH5a and a few of the
resulting white Ap^R colonies that appeared on the LB-agar plates
containing X-gal and ampicillin were analyzed for their plasmid content.
Colonies carrying pABX95DN exhibited fragments of 8.8 and 7.2 kb visible
in agarose gels after EcoRI and HindIII digestion.

Example 7
Construction of Sac. erythraea AKR5 carrying the eryAKR5 allele by gene replacement

Approximately 1 mg of plasmid pABX95DN, isolated from E. coli
K12 DH5α/pABX95DN, was transformed into Sac. erythraea NRRL 2338
and stable Th^R colonies were isolated. Serial dilutions of one of these
colonies were screened for the loss of the antibiotic resistance marker and
total DNA from 5 Th^S colonies as well as from untransformed Sac. ·
erythraea NRRL 2338 was digested with SstI and analyzed by Southern

hybridization employing the 0.8 kb SalI fragment between sequence
coordinates 24.26 and 25.06 (from pABX9) as probe. Whereas NRRL 2338
showed one SstI band of 3.7 kb that hybridized to the probe, samples from
four of the Th^S strains exhibited a SstI-hybridizing band of 6.1 kb

5   indicating the presence of the mutant allele. One of these colonies was
kept and designated strain AKR5. It carries a deletion of 813 bp in the KR5
segment of eryA and is designated the eryAKR5 allele.

Example 8

Isolation, purification and properties of 5-oxo-5,6-dideoxy-3-a-mycarosyl

10   erythronolide B from Sac. erythraea AKR5

A 10-liter fermentation of Sac. erythrea AKR5 carrying the eryAKR5
allele in a Biolafitte fermentor using SNC Media. The fermentor was

15   inoculated with 100 ml of a 3 day old seed. The pO2 was initially 80 ppm
and the temperature was maintained at 32°C. The pH was controlled to
7.0 ± 0.2 by addition of propionic acid or potassium hydroxide as needed.
At harvest (3 days), the whole broth was extracted three times with 4-liter
portions of ethylacetate. The combined extracts were concentrated under

20   reduced pressure and the residue was chromatographed on a column (50 x
5 cm) of silica gel packed and loaded in toluene and eluted with a stepwise
gradient of increasing concentration of isopropanol in toluene. Fractions
were analyzed by TLC and spots were detected by spraying with
anisaldehyde sulfuric acid spray reagent and heating. A major component

25   giving blue colored spots eluted with approximately 7% isopropanol.
Fractions containing this material were combined and concentrated to a
residue (800 mg). This was further chromatographed on a column (100 x
3 cm) of Sephadex LH-20 in chloroform-heptane-ethanol, 10:10:1, v/v/v.
Fractions were analyzed as above, early fractions (9-13) yielded 5,6-dideoxy-

30   3-a-mycarosyl-5-oxoerythronolide B (45 mg) which was crystallized from
heptane/ethylacetate mixture to mp 163-164 °C.

15

CMR spectrum in CDCl₃ (ppm downfield from TMS)

8.6	37.9	70.0
9.9	38.7	76.2
9.9	40.4	76.4
10.4	40.7	80.4
14.5	43.3	100.4
15.2	45.8	175.8
17.1	46.8	210.8
17.7	48.9	217.7
25.3	66.5	
25.5	69.4	

Structure was determined by single crystal X-ray diffraction.

5   Later fractions (15-17) yielded 5,6-dideoxy-5-oxoerythronolide B
(10 mg) and still later fractions yielded 5,6-dideoxy-6,6a-epoxy-5-
oxoerythronolide B (2.8 mg).

10   Example 9
                    Construction of plasmid pALeryAKR2

       The 1.3 kb DNA segment comprised between coordinates 8.63-9.93
(fragment 1) is amplified by PCR employing two oligodeoxynucleotides, 1a
15   (5'-GGGAGCATGCTCTCGGTGCGCGGCGGCCGC-3') and 1b (5'-
GCCCTGCAGCGCGTACTCCGAGGTGGCGGT-3'). Similarly, the 1.3 kb
DNA segment between coordinates 9.99-11.26 (fragment 2) is PCR-
amplified employing primers 2a (5'-
TGGTCTGCAGGCGAGGCCGGACACCGAGG-3') and 2b (5'-
20   GGAAGAAGTCAAAGTTCCTCGGTCCCTTCT-3'). After digestion with
SphI + PstI (fragment 1) and PstI + EcoRI (fragment 2), the two fragments
are ligated to EcoRI + SphI-digested pWHM3. The resultant mixture
contains the desired plasmid pALeryAKR2.

Example 10
Construction of E. coli K12 DH5a/pALeryAKR2

Approximately 10 ng of plasmid pALeryAKR2, prepared as described in Example 9, are transformed into E. coli K12 DH5α, and a few of the resulting white ApR colonies that appear on the LB-agar plates containing X-gal and ampicillin are analyzed for their plasmid content. The identity of plasmid pALeryA2KR2, 9.8 kb in size, and carrying a 2.6 kb EcoRI-SphI insert with an internal PstI site, is verified by SalI digestion (fragments at 2.91, 2.21, 1.61, 1.42, 1.08, 0.29, 0.12 and 0.10 kb are released, visible by agarose gel electrophoresis). pALeryAKR2 contains an in-frame deletion of 102 base pairs of the corresponding segment of the wild type eryA chromosomal DNA. The cloned segment in pALeryAKR2 is designated the eryAKR2 allele.

Example 11
Construction of Sac. erythraea AKR2 carrying the eryAKR2 allele by gene replacement

Approximately 1 mg of plasmid pALeryAKR2, isolated from E. coli K12 DH5α/pALeryAKR2, is transformed into Sac. erythraea protoplasts and stable ThR colonies are isolated. Serial dilutions of one of these colonies are screened for loss of the antibiotic resistance marker, and six ThS colonies are analyzed for their genotype by Southern hybridization. Total DNA from the six ThS colonies and from untransformed Sac. erythraea NRRL2338 is digested with PstI and with SalI and is then examined by Southern hybridization using the 2.6 kb EcoRI-SphI insert from pALeryAKR2 as probe. Whereas NRRL2338 contains a 39 kb PstI hybridizing band, colonies in which the mutation in KR2 has been introduced (strain AKR2) exhibit two bands of approximately equal intensity, one at 27 kb and the other at 12 kb. The SalI digest, with bands at 1.04, 0.75, 0.29, 0.12 and 0.10 kb common to NRRL2338 and AKR2, but with the 1.16 kb band in NRRL2338 replaced by the 1.06 kb band in AKR2, confirms that the only change introduced into strain AKR2 is the deletion of the 102 bp segment from KR2, resulting in a strain carrying the eryAKR2 allele.

Example 12
Isolation and purification of 11-deoxy-11-oxoerythromycin A

5   The fermentation beer of strain AKR2, cooled to 4°C is adjusted to pH 8.0 and is extracted sequentially with three equal volumes of methylene chloride. The combined methylene extracts are concentrated to an oily residue and partitioned between heptane and methanol. The methanol layer is removed, washed once with heptane and concentrated to a residue. The residue is digested in methylene chloride and washed

10   once with potassium phosphate buffer pH 7.8 and once with water. The methylene chloride layer is concentrated to a residue and digested in the lower phase (1:1:1, v/v/v) of a carbon tetrachloride; methanol; aqueous phosphate buffer (0.05 M, pH 7.0) system and chromatographed on an Ito Coil Planet Centrifuge in the same system. Fractions containing the

15   desired 11-oxo-11-deoxyerythromycin A were combined, concentrated, digested in methylene chloride, washed well with water and concentrated on rotary evaporator under reduced pressure to yield 11-deoxy-11-oxoerythromycin A as an off-white solid froth. Its identity is confirmed by comparison with antibiotic L53-18A. 11-Deoxy-11-oxoerythromycin A is

20   dissolved in tetrahydrofuran and the solution is diluted with an equal volume of water. This is then acidified to pH 4.0 and allowed to stand at room temperature for 4 hours. The pH is adjusted to 9.0 and the solution is diluted with an equal volume of water and extracted with two volumes of methylene chloride. The combined methylene chloride extracts were

25   evaporated to dryness under reduced pressure to yield antibiotic L53-18A as a white solid.

Example 13
Construction of plasmid pALeryADH4

30

    Primers 3a (GCGCGAGCTCGACGACCAGGGCGGCATGGT) and 3b (GGTGGCATGCTGCGACCACTGCGCGTCGGC) are used to PCR-amplify the 1.05 kb *eryA* segment of the Sac. erythraea chromosome between sequence coordinates 18.47-20.07 (fragment 3), and primers 4a

35   (AGCTGCATGCTCTGGACTGGGGACGGCTAG) and 4b (CGCGGGATCCCAGCTCCCACGCCGATACCG) are used to amplify the 1.35 kb segment between sequence coordinates 20.58-21.96 (fragment 4) as described in Example 1. Fragment 3 and 4, after digestion with SstI + SphI

and with SphI + BamHI, respectively, are ligated to SstI -, BamHI-digested
pWHM3. The resulting ligation mixture contains the desired plasmid
pALeryADH4.

5   Example 14
                    Construction of E. coli K12 DH5a/pALeryADH4

            Approximately 10 ng of pALeryADH4, prepared as described in
        Example 13, are transformed transformed into E. coli K12 DH5α, and a few
10   of the resulting white Ap^R colonies that appear on the LB-agar plates
        containing X-gal and ampicillin are analyzed for their plasmid content.
        The identity of plasmid pALeryADH4, 9.6 kb in size, is verified by SphI +
        EcoRI digestion (fragments at 7.2, 1.35 and 1.05 kb are released).
        pALeryADH4 carries a 498 base pair in-frame deletion of the
15   corresponding segment of the wild type eryA DNA. The cloned segment
        in pALeryADH4 is designated the eryADH4 allele.

                                     Example 15
        Construction of Sac. erythraea ADH4 carrying the eryADH4 allele by gene
20   replacement

            Approximately 1 mg of plasmid pALeryADH4, isolated from E. coli
        K12 DH5α/pALeryADH4, is used for transformation into Sac. erythraea
        protoplasts and stable Th^R colonies are isolated. Serial dilutions of one of
25   these colonies are screened for loss of the antibiotic resistance marker, and
        six Th^S colonies are analyzed for their genotype by Southern
        hybridization. Total DNA from the six Th^S colonies and from
        untransformed Sac. erythraea NRRL2338 is digested with SphI and with
        SstI and examined by Southern hybridization using the 2.4 kb SstI-BamHI
30   insert from pALeryADH4 as probe. Strains in which the wild type allele
        has been replaced by the mutated copy show two SphI bands, one at 13.5 kb
        and the other at 12.4 kb, whereas the wild type strain exhibits a single band
        at 26 kb. The SstI pattern, with the 2.9 kb band from NRRL2338 being
        replaced in ADH4 by a 2.5 kb band, confirms that the 487 bp deletion
35   created in plasmid pALeryADH4 has been transferred into the
        chromosome of ADH4. Strains that carry the eryADH4 allele in place of
        the wild type sequence are designated Sac. erythraea ADH4.

Example 16

Isolation and characterization of 7-hydroxyerythromycin A and 6-deoxy-7-hydroxyerythromycin A

5   The fermentation beer of strain ADH4 is cooled to 4°C and the pH is adjusted to 5.0. The mixture is extracted once with an equal volume of methylene chloride. The pH of the aqueous layer is readjusted to 9.0 and two further methylene chloride extracts are carried out. These two extracts are combined, washed with water and concentrated to a residue. This is

10   digested in 10 ml of the upper phase of a (3:7:5, v/v/v) mixture of hexane, ethylacetate, aqueous phosphate buffer (0.05 M, pH 7.5) and chromatographed on an Ito Coil Planet Centrifuge in the same system. Fractions containing the desired 7-hydroxyerythromycin were combined, concentrated, and partitioned between methylene chloride and dilute (pH

15   9.5) ammonium hydroxide solution. Fractions containing the desired 6-deoxy-7-hydroxyerythromycin were combined, concentrated, and partitioned between methylene chloride and dilute (pH 9.5) ammonium hydroxide solution. The methylene chloride layers are washed with water and then concentrated to yield the desired 7-hydroxyerythromycin A and

20   6-deoxy-7-hydroxyerythromycin A as white foams.

Example 17

Construction of plasmid pALeryAKS1

25   The 1.4 kb segment of *eryA* , between sequence coordinates 1.11-2.54 (fragment 5) and the 1.5 kb segment between sequence coordinates 2.88-4.37 (fragment 6) are PCR-amplified using primers 5a (TGCAGAATTCGCTGGCCGCGCTCTGGCGCT) and 5b (GAGAGCTGCAGCATGAGCCGCTGCTGCGGG), and 6a

30   (CATGCTGCAGGACTTCAGCCGGATGAACTC) and 6b (GAGGAAGCTTCCAGCCGGTCCAGTTCGTCC), respectively, as described in Example 9. After digestion with EcoRI + PstI (fragment 5) and PstI + HindIII (fragment 6), the two fragments are ligated to EcoRI + HindIII-digested pWHM3. The resulting mixture contains the desired plasmid

35   pALeryAKS1.

20

Example 18
Construction of E. coli K12 DH5a/pALeryAKS1

5   Approximately 10 ng of pALeryAKS1, prepared as described in Example 17, are transformed into E. coli K12 DH5α, and a few of the resulting white Ap^R colonies that appear on the LB-agar plates containing X-gal and ampicillin are analyzed for their plasmid content. The identity of plasmid pALeryAKS1, 10.1 kb in size, is verified by digestion with PstI + HindIII (fragments of 8.6 and 1.5 kb are observed by agarose gel

10   electrophoresis) and with SalI (fragments of 2.93, 2.21, 1.42, 1.37, 0.86, 0.54, 0.27, 0.14, 0.13, and 0.10 kb are observed). pALeryAKS1 carries an in-frame deletion of 282 base pairs of the corresponding wild type eryA DNA. The cloned insert in plasmid pALeryAKS1 is designated the eryAKS1 allele.

15   Example 19
Construction of Sac. erythraea AKS1 carrying the eryAKS1 allele by gene
replacement

      Approximately 1 mg of plasmid pALeryAKS1, isolated from E. coli

20   K12 DH5α/pALeryAKS1, is used for transformation into Sac. erythraea protoplasts and stable Th^R colonies are isolated. Serial dilutions of one of these colonies are screened for loss of the antibiotic resistance marker, and six Th^S colonies are analyzed for their genotype by Southern hybridization. Total DNA from the six Th^S colonies and from

25   untransformed Sac. erythraea NRRL2338 is digested with PstI and with SmaI and examined in Southern hybridization employing the 2.9 kb EcoRI-HindII insert from pALeryAKS1 as probe. Colonies in which the wild type allele has been replaced by the mutated copy (strain AKS1) show two PstI bands, one at 34.5 and the other at 4.4 kb, whereas the wild type

30   strain exhibits a single band at 39 kb. The SmaI pattern, with the 2.9 kb band from NRRL2338 being replaced in AKS1 by a 2.6 kb band, confirms that the 282 bp created in plasmid pALeryAKS1 has been transferred into strain AKS1. Strains that carry the eryAKS1 allele are designated Sac. erythraea AKS1.

35

Example 20
Synthesis of (2S,3R,4S,5S)3,5-dihydroxy-2,4-dimethylhexanoic acid n-butyl
thioester

5   A convenient source of this compound in chiral purity is the
antibiotic oleandomycin. Oleandomycin (5 g) is dissolved in an aprotic
solvent such as toluene and treated with diazabicyclo[5.4.0]undecene-5
(1 g) and heated for one hour. The resulting solution is poured into iced
water, agitated well and the organic layer is drawn off and concentrated to
10   a residue. The residue is digested in methylene chloride and treated
exhaustively with a solution of ozone. The resulting ozonide is
oxidatively decomposed with dilute hydrogen peroxide in sufficient
aqueous ethanol to yield a monophasic mixture. This is further diluted
with water and made 0.1 N with sodium hydroxide. The mixture is
15   warmed for one hour at 70°C and then cooled before being acidified to
pH 2.5 with dilute sulfuric acid. The mixture is then exhaustively
extracted with methylene chloride. The combined extracts are
concentrated to an oily residue and the desired lactone is recovered by
chromatography on silica gel eluted with a gradient of toluene-
20   isopropanol.

        The δ-lactone is converted to the butyl thioester before feeding to
Sac. erythrea AKS1 by refluxing with n-butylthiol in the presence of a
catalytic amount of triethylamine.

25   ## Example 21
Isolation of (14S,15S)14(1-hydroxyethyl)erythromycin A

        The fermentation broth of AKS1 is cooled to 4°C and adjusted to pH
4.0 and extracted once with methylene chloride. The aqueous layer is
30   readjusted to pH 9.0 and extracted twice with methylene chloride and the
combined extracts are concentrated to a solid residue. This is digested in
methanol and chromatographed over a column of Sephadex LH-20 in
methanol. Fractions are tested for bioactivity against a sensitive organism,
such as Staphylococcus aureus Th^R, and active fractions are combined.
35   The combined fractions are concentrated and the residue is digested in 10
ml of the upper phase of a solvent system consisting of n-heptane,
benzene, acetone, isopropanol, 0.05 M, pH 7.0 aqueous phosphate buffer
(5:10:3:2:5, v/v/v/v/v), and chromatographed on an Ito Coil Planet

Centrifuge in the same system. Active fractions are combined,
concentrated and partitioned between methylene chloride and dilute
ammonium hydroxide (pH 9.0). The methylene chloride layer is
separated and concentrated to yield the desired product as a white foam.

Example 22
Construction of plasmid pALeryAKS2

Primers 7a (CGCCCGAATTCGAGGCGCTGGGCGCCCGGAC) and 7b
(CCACCTGCAGCGCGGGACCTTCCAGCCCC), and primers 8a
(GTGGGTCGCTGCAGACGGTGACTGCGG) and 8b
(GGTCAAGCTTCGTCGGCGAGCAGCTTCTC) are used to PCR-amplify
the 1.45 kb *eryA* segment between sequence coordinates 5.71-7.16
(fragment 7) and the 1.5 kb *eryA* segment between sequence coordinates
7.22-8.70 (fragment 8), respectively. After digestion with EcoRI + PstI
(fragment 7) and with PstI + HindIII (fragment 8), the two fragments are
ligated to pWHM3 cut with EcoRI + HindIII. The resulting mixture
contains the desired plasmid pALeryAKS2.

Example 23
Construction of E. coli K12 DH5a/pALeryAKS2

Approximately 10 ng of pALeryAKS2, prepared as described in
Example 22, are transformed into E. coli K12 DH5α, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAKS2, 10.1 kb in size, is verified by digestion with PstI
+ HindIII (fragments of 8.6 and 1.5 kb are observed by agarose gel
electrophoresis) and with SstII (fragments of 4.0, 2.3, 2.0, 0.72, 0.43, 0.40,
0.20, 0.18, 0.13 and 0.11 kb observed). Plasmid pALeryAKS2 carries an in-
frame deletion of 60 base pairs of the corresponding wild type *eryA* DNA.
This deletion removes the active site cysteine from KS2. The cloned
insert in plasmid pALeryAKS2 is designated the eryAKS2 allele.

Example 24

Construction of Sac. erythraea AKS2 carrying the eryAKS2 allele by gene replacement

5   Approximately 1 mg of plasmid pALeryAKS2, isolated from E. coli
K12 DH5α/pALeryAKS2, is used for transformation into Sac. erythraea
protoplasts and stable Th^R colonies are isolated. Serial dilutions of one of
these colonies are screened for loss of the antibiotic resistance marker, and
six Th^S colonies are analyzed for their genotype by Southern
10   hybridization. Total DNA from the six Th^S colonies and from
untransformed Sac. erythraea NRRL2338 is digested with PstI and with
SstII and examined in Southern hybridization employing the 2.9 kb EcoRI-
HindII insert from pALeryAKS2 as probe. Colonies in which the wild type
allele has been replaced by the mutated copy (strain AKS2) show two PstI
15   bands, one at 34.5 and the other at 4.4 kb, whereas the wild type strain
exhibits a single band at 39 kb. The SstII pattern, with the 0.78 kb band
from NRRL2338 being replaced in AKS2 by a 0.72 kb band, confirms that
the 60 bp created in plasmid pALeryAKS2 has been transferred into strain
AKS2. Strains that carry the eryAKS2 allele are designated Sac. erythraea
20   AKS2.

Example 25

Synthesis of (2R,3R,4S,5R)2,4-dimethyl-3-fluoro-5-hydroxyhexanoic acid n-butyl thioester

25

(2R,3S,4S,5R)3,5-Dihydroxy-2,4-dimethylhexanoic acid-δ-lactone (1
g) from Example 20 is digested in 10 ml of pyridine and treated with p-
toluenesulfonyl chloride (1.3 g) and allowed to stand at room temperature
overnight. The mixture is poured into iced water and extracted with
30   methylene chloride and the methylene chloride is concentrated to the
crude sulfonate ester. This is digested in acetonitrile (100 ml) and heated
under reflux after the addition of tetrabutylammonium fluoride (1.75 g).
After 6 hours the mixture is cooled, poured over iced water (300 ml) and
extracted three times with 200 ml portions of methylene chloride. The
35   combined methylene chloride extracts were concentrated and the residue
was chromatographed on a column of silica gel eluted with a stepwise
gradient of isopropanol (0 to 50%) in toluene. Fractions containing
(2R,3R,4S,5R)2,4-dimethyl-3-fluoro-5-hydroxyhexanoic acid d-lactone were

combined and concentrated to a white solid. The lactone is then converted to the n-butyl thiolester by refluxing in n-butyl thiol with a catalytic amount of triethylamine. Solvent is removed and the residue is digested in DMSO before feeding to fermentations of Sac. erythraea AKS2.

5

Example 26
Isolation and purification of 11-epifluoro-15-norerythromycin A

The fermentation broth of strain AKS2 is cooled to 4°C and adjusted

10   to pH 4.0 and extracted once with ethylacetate. The aqueous layer is adjusted to pH 9.0 and extracted twice with methylene chloride and the combined extracts are concentrated to a white solid. This is chromatographed over a column of Sephadex LH-20 in a mixture of heptane, chloroform, ethanol (10:10:1, v/v/v) and fractions containing the

15   desired product are combined and concentrated to a solid residue. This is further purified by countercurrent chromatography on an Ito Coil Planet Centrifuge on a system composed of carbon tetrachloride; methanol; 0.05 M; pH 7.0 aqueous potassium phosphate buffer (1:1:1, v/v/v). Fractions containing the desired 11-epifluoro-15-norerythromycin were combined,

20   and concentrated to a residue. This was digested in methylene chloride and dilute (pH 9.5) ammonium hydroxide and the methylene chloride layer was separated, washed with water and concentrated to yield the desired 11-epifluoro-15-norerythromycin A as white solid.

25   ## Example 27
Construction of plasmid pALeryAM4.1

Primers 9a (GCGCCGAATTCTCGAGACGGCGTGGGAGGCA) and 9b (TTGCGGTACCAGTAGGAGGCGTCCATCGCG) are employed to PCR-

30   amplify the 2.0 kb eryA segment between sequence coordinates 17.35-19.38 (fragment 9). After digestion with EcoRI + KpnI, fragment 9 is ligated to pUC19 cut with the same two enzymes The resulting mixture contains the desired plasmid pALeryAM4.1.

25

Example 28
Construction of E. coli K12 DH5a/pALeryAM4.1

Approximately 10 ng of pALeryAM4.1, prepared as described in
5   Example 27, are transformed into E. coli K12 DH5a, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAM4.1, 4.7 kb in size, is verified by digestion with SalI
(fragments of 2.8, 0.85, 0.53, 0.27 and 0.22 kb are observed by agarose gel
10   electrophoresis).

Example 29
Construction of plasmid pALeryAM4.2

15   Primers 10a (GCTGGGATCCCGCGGCGCGGGTTGCAGCAC) and
10b (CGGAACTCGGTGAGCATGCCGGGACTGCTC) are used to PCR-
amplify the 2.1 kb eryA segment between sequence coordinates 21.94-24.00
(fragment 10). The 2.6 kb fragment KpnI(96)-BamHI(102) from cosmid
clone pR1, and fragment 10 cut with BamHI + SphI, are ligated to
20   pALeryAM4.1 cut with KpnI + SphI. The resulting mixture contains the
desired plasmid pALeryAM4.2.

Example 30
Construction of E. coli K12 DH5a/pALeryAM4.2

25

Approximately 10 ng of pALeryAM4.2, prepared as described in
Example 29, are transformed into E. coli K12 DH5a, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
30   of plasmid pALeryAM4.2, 9.3 kb in size, is verified by digestion with XhoI
+ SphI (to ensure that the entire 6.65 kb insert is released) and with SalI,
with fragments of 2.8, 1.82, 1.09, 0.94, 0.85, 0.75, 0.45, 0.27, 0.22 and 0.13 kb
are observed by agarose gel electrophoresis).

Example 31
Construction of plasmid pALeryAM1

The 2.9 kb SmaI(4)-SmaI(20) fragment from cosmid clone pR1 is
ligated to pUC12 cut with SmaI. The resulting mixture contains plasmid
pALeryAM1.

Example 32
Construction of E. coli K12 DH5αa/pALeryAM1

Approximately 10 ng of pALeryAM1, prepared as described in
Example 31, are transformed into E. coli K12 DH5α, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAM1, 5.6 kb in size, is verified by digestion with SmaI
(the 2.9 kb insert is realeased) and with SphI, with release of one 4.4 and
one 1.07 kb bands. Both orientations of the insert in plasmid pALeryAM1
are useful.

Example 33
Construction of plasmid pALeryAM4.3

Plasmid pALeryAM1 is cut with XhoI to completion, partially with
SphI, and the resulting 5.25 kb band, isolated from an agarose gel, is ligated
to the 6.65 kb insert released from pALeryAM4.2 by XhoI + SphI digestion
The resulting mixture contains the desired plasmid pALeryAM4.3.

Example 34
Construction of E. coli K12 DH5a/pALeryAM4.3

Approximately 10 ng of pALeryAM4.3, prepared as described in
Example 33, are transformed into E. coli K12 DH5α, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAM4.1, 11.9 kb in size, is verified by XhoI + SphI
digestion (fragments of 6.65 and 5.25 kb are visible by agarose gel-
electrophoresis). Plasmid pALeryAM4.3 carries the entire eryA module 4

27

inserted into the KS region of module 1. The cloned insert in
pALeryAM4.3 is degnated the eryAM412 allele.

Example 35
5   Construction of plasmid pALeryAM4.4

Plasmid pALeryAM4.3 is cut with EcoRI + HindIII, and the
resulting 9.2 kb band, recovered from an agarose gel, is ligated to pWHM4
cut with the same two enzymes. The resulting mixture contains the
10   desired plasmid pALeryAM4.4.

Example 36
Construction of E. coli K12 DH5α/pALeryAM4.4

15   Approximately 10 ng of pALeryAM4.4, prepared as described in
Example 35, are transformed into E. coli K12 DH5α, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAM4.1, 16.5 kb in size, is verified by EcoRI + HindIII
20   digestion, with fragments of 9.2 and 7.3 kb released. Plasmid
pALeryAM4.4 carries the eryAM412 allele on the Sac. erythraea multicopy
vector pWHM4.

Example 37
25   Construction of Sac. erythraea AM412 carrying the eryAM412 allele by
gene conversion

Approximately 1 mg of plasmid pALeryAM4.4, isolated from E. coli
K12 DH5α/pALeryAM4.4, is used for transformation into Sac. erythraea
30   strain AKS1 protoplasts. A few hundred transformants are screened for
antibiotic production by the agar-plug assay, and one of the colonies found
to produce antimicrobial activity is cured of pALeryAM4.4 by protoplast
formation and regeneration as described in General Methods. Total DNA
from six antibiotic-producing, Th^S colonies (strain AM412)and from strain
35   AKS1 is digested with SphI and with XhoI and the resulting Southern blot
is hybridized first to the 2.9 kb insert from pALeryAM1, and then to the 2.9
kb SstI(95)-SstI(101) fragment from plasmid pALeryAM4.2. With the first
probe, the SphI band at 0.8 kb in strain AKS1 is seen to be replaced by a 7.5

kb band in strain AM412, whereas the other two bands at 2.4 kb and 5.2 kb are unaffected. In the XhoI digest, the AKS1 band at 2.9 kb is replaced by a 9.6 kb band in AM412, with the other band at 5.2 kb conserved in both strains. Using the SstI(95)-SstI(101) fragment as probe, strain AKS1 exhibits one band at 25.5 kb and one at 17.9 kb in the SphI and XhoI digests, respectively, whereas, in addition to these bands, strain AM412 shows one SphI band at 7.5 kb and one XhoI band at 9.6 kb. In this way, it is established that the eryAKS1 allele has been converted into the eryAM412 allele in strain AM412.

Example 38
Isolation and purification of 14-(1-propyl)erythromycin A

At harvest the fermentation is adjusted to pH 9.5 and extracted twice with equal volumes of methylene chloride. The combined extracts are washed once with water and concentrated to an oily residue. This is partitioned in a heptane methanol water (5:5:1, v/v/v) system and the lower layer is washed once with heptane and then concentrated to a semisolid residue. This is digested in methanol and chromatographed over a column of Sephadex LH-20 in methanol. Fractions are tested for bioactivity in an agar diffusion assay on plates seeded with the macrolide-sensitive strain Staphylococcus aureus Th^R. Active fractions are combined and further purified by chromatography over silica gel a chloroform:methanol gradient containing 0.1% triethylamine. Fractions containing the desired 14-(1-propyl)erythromycin A are combined and concentrated to yield the product as a white solid.

Example 39
Construction of plasmid pALeryAM5.1

The 4.7 kb eryA fragment between sequence coordinates 23.65-28.36 (fragment 11) is PCR-amplified employing primers 11a (ATGCTCGAGATCTCGTGGGAGGCGCTGGA) and 11b (AGAACTCGGTGAGCATGCCCGGGCCCGCCA). Fragment 11, after digestion with XhoI + SphI, is ligated to the 5.25 kb fragment resulting from complete XhoI and partial SphI digestion of pALeryAM1, as in Example 33. The resulting mixture contains the desired plasmid pALeryAM5.1.

29

Example 40
Construction of E. coli K12 DH5α/pALeryAM5.1

5   Approximately 10 ng of pALeryAM5.1, prepared as described in
Example 39, are transformed into E. coli K12 DH5α, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAM5.1, 9.95 kb in size, is verified by SphI + XhoI
10   digestion, with fragments of 5.25 and 4.7 kb released, and by SmaI
digestion where fragments of 3.39, 2.68 and 1.94 (doublet) kb are observed.
Plasmid pALeryAM5.1 carries the entire eryA module 5 inserted into the
β–ketoacyl ACP synthase region of module1. The cloned insert in plasmid
pALeryAM5.1 is designated the eryA512 allele.

15

Example 41
Construction of plasmid pALeryAM5.2

       Plasmid pALeryAM5.1 is cut with EcoRI + HindIII and the resulting
20   6.3 kb fragment, recovered from an agarose gel, is ligated to pWHM4 cut
with the same two enzymes. The resulting mixture contains the desired
plasmid pALeryAM5.2.

Example 42
25   ## Construction of E. coli K12 DH5α/pALeryAM5.2

       Approximately 10 ng of pALeryAM5.2, prepared as described in
Example 41, are transformed into E. coli K12 DH5α, and a few of the
resulting white Ap^R colonies that appear on the LB-agar plates containing
30   X-gal and ampicillin are analyzed for their plasmid content. The identity
of plasmid pALeryAM5.2, 13.6 kb in size, is verified by digestion with
EcoRI + HindIII, with fragments of 7.3 and 6.3 kb released. Plasmid
pALeryAM5.2 contains the eryAM512 allele on the Sac. erythraea
multicopy vector pWHM4.

35

Example 43
Construction of Sac. erythraea AM512 carrying the eryAM512 allele by gene conversion

5   Approximately 1 mg of plasmid pALeryAM5.2, isolated from E. coli
K12 DH5α/pALeryAM5.2, is used for transformation into Sac. erythraea
strain AKS1 protoplasts. A few hundred transformants are screened for
antibiotic production by the agar-plug assay, and one of the colonies found
to produce antimicrobial activity is cured of pALeryAM5.2 by protoplast

10   formation and regeneration as described in General Methods. Total DNA
from six antibiotic-producing, Th^S colonies (strain AM512)and from strain
AKS1 is digested with SphI and with XhoI and the resulting Southern blot
is hybridized first to the 2.9 kb insert from pALeryAM1, and then to the 0.8
kb NcoI(119)-NcoI(123) fragment from plasmid pALeryAM5.1. With the

15   first probe, the SphI band at 0.8 kb in strain AKS1 is replaced by a 5.5 kb
band in strain AM512, whereas the other two bands at 2.4 kb and 5.2 kb are
unaffected. In the XhoI digest, the AKS1 band at 2.9 kb is replaced by a 7.6
kb band in AM512, with the other band at 5.2 kb conserved in both strains.
Using the NcoI(119)-NcoI(123) fragment as probe, strain AKS1 exhibits one

20   band at 25.5 kb and one at 17.9 kb in the SphI and XhoI digests,
respectively, whereas, in addition to these bands, strain AM512 shows one
SphI band at 5.5 kb and one XhoI band at 7.6 kb. In this way, it is
established that the eryAKS1 allele has been converted into the eryAM512
allele in strain AM512.

25

Example 44
Isolation and purification of 14[1(1-hydroxypropyl)]erythromycin A

         At harvest the pH of the fermentation of AM512 is adjusted to 9.5
30   and the mixture is extracted twice with equal volumes of ethylacetate.
The combined ethylacetate extracts are washed with water, dried and
partitioned in a heptane, methanol, water (5:5:1, v/v/v) system. The
lower (methanolic phase) is washed with an equal volume of heptane and
is concentrated to a residue. This is chromatographed on a Sephadex LH-

35   20 column in methanol and fractions containing the desired 14[1(1-
hydroxypropyl)]erythromycin A are concentrated and further purified by
chromatography on an Ito Coil Planet Centrifuge in a system consisting of
n-heptane, benzene, acetone, isopropanol, 0.65 M, pH 7.0 aqueous

potassium phosphate buffer (5:10:2:3:5, v/v/v/v/v). Fractions containing
the desired product are concentrated to a solid residue and partitioned
between methylene chlorid and dilute (pH 9.5) ammonium hydroxide.
The organic layer is washed with water and concentrated to yield 14[1(1-
hydroxypropyl)]erythromycin A as a white solid.

Although the present invention is described in the Examples listed
above in terms of preferred embodiments, they are not to be regarded as
limiting the scope of the invention. The above descriptions serve to
illustrate the principles and methodologies involved in creating the three
types of mutations that can be introduced into the *eryA* segment of the
Sac. erythraea chromosome that result in the synthesis of novel
polyketide products. Although single Type I alterations, leading to the
production of 5-oxo-5,6,-dideoxy-3α-mycorosyl erythronolide B, 11-oxo-11-
deoxyerythromycin A, 7-hydroxyerythromycin A, 7-oxo-
7deoxyerythromycin A, 5-desosaminyl-3-oxo-3-deoxyerythronolide A, and
Δ-6,7-anhydro-6-deoxyerythromycin A are specified herein, it is obvious
that other Type I changes can be introduced into the *eryA* segment
leading to novel polyketide structures. Among the additional Type I
alterations that can be obtained are those in which two or more modules
are affected leading to the synthesis of novel polyketides. Examples of
combinations of two Type I alterations leading to useful compounds
include but are not limited to: mutants of the the β-ketoreductase of
module 2 (KR2) and the β-ketoreductase of module 4 (KR4) leading to the
formation of 7,11-dioxo-7,11-dideoxyerythromycin A; mutants of KR2 and
the β-ketoreductase of module 6 (KR6) leading to the formation of 3,11-
dioxo-3,11-dideoxy-5-desosaminylerythronolide A; mutants of KR2 and
the dehydratase of module 4 (DH4) leading to the synthesis of 7-hydroxy-
11-oxo-11-deoxyerythromycin A; mutants of KR2 and the enoylreductase
of module 4 (ER4) leading to the synthesis of Δ-6,7-anhydro-11-oxo-11-
deoxyerythromycin A; mutants of KR4 and KR6 leading to the synthesis of
3,7-dioxo-3,7-dideoxy-5-desosaminylerythronolide A; mutants of KR6 and
DH4 leading to the synthesis of 3-oxo-3-deoxy-5-desosaminyl-7-
hydroxyerythronolide A; mutants of KR6 and ER4 leading to the synthesis
of 3-oxo-3-deoxy-5-desosaminyl-Δ-6,7-anhydroerythronolide A. Examples
of combinations of three Type I alterations leading to the synthesis f
novel polyketides include but are not limited to: mutants of KR2, KR4 and
KR6 leading to the synthesis of 3,7,11-trioxo-3,7,11-trideoxy-5-

desosaminylerythronolide A; mutants of KR2, KR6 and DH4 leading to
the synthesis of 3,11-dioxo-3,11-dideoxy-5-desosaminyl-7-
hydroxyerythronolide A; mutants of KR2, KR6 and ER4 leading to the
synthesis of 3,11-dioxo-3,11-dideoxy-5-desosaminyl-D-6,7-

5   anhydroerythronolide A. All combinations of two or three Type I
mutants, the Sac. erythraea strains that carry said combinations and the
corresponding polyketides produced from said strains, therefore, are
included within the scope of the present invention.

         Although the Type II mutants specified herein have been
10   constructed in the β-ketoacyl ACP synthase of module 1 (KS1) and the β-
ketoacyl ACP synthase of module 2 (KS2), other Type II mutants can be
constructed in other domains to result in the synthesis of novel
polyketide structures upon feeding with appropriate substrate analogs.
Other Type II mutants include but are not limited to: inactivation of the
15   either of the acyltransferases or acyl carrier proteins of module 1, or the
acyltransferase or acyl carrier protein of module 2, the β-ketoacyl ACP
synthase, acyltransferase or acyl carrier protein of module 3, module 4 or
module 5. Furthermore, compounds other than (2S,3R,4S,5S)3,5-
dihydroxy-2,4-dimethylhexanoic acid-ethyl thioester and (2S,3S,4S,5S)2,4-
20   dimethyl-3-fluoro-5-hydroxyhexanoic acid-ethyl thioester specified herein
can be synthesized and fed to strains AKS1 or AKS2 specified herein or
other strains that carry other Type II mutations to result in the creation of
novel polyketides that are within the scope of the present invention.

         Although two examples of Type III alterations are specified herein,
25   it is apparent to those skilled in the art that many other examples of Type
III changes are possible. Strains of Sac. erythraea carrying changes of this
type offer the very high potential for the production of novel polyketides
of specified structure, since they do not require synthetic substrates as do
Type II mutants and they are not limited to the formation of derivatives
30   of erythromycin, as in the case of Type I mutants. In the embodiments of
Type III mutants specified herein, we have illustrated how a second copy
of a complete module can be introduced at a desired position by gene
conversion to result in the synthesis of 14-(1-propyl)erythromycin A or 14-
[1(1-hydroxypropyl])erythromycin A. These alterations make use of the
35   high conservation and simultaneous lack of specificity of the β-ketoacyl
ACP synthases of modules 1 and 2, thereby making possible the
construction of hybrid β-ketoacyl ACP synthase functions consisting of
portions of proteins derived from different modules. Those skilled in the

art understand, therefore, that it is possible, as exemplified for KS1 and
KS2, to delete a small portion of the β–ketoacyl ACP synthase of other
modules and to construct strains carrying such alterations which can then
be employed as hosts for introducing at the deleted β–ketoacyl ACP

5   synthase location a second copy of any homologous module.
Furthermore, as exemplified herein, it is also possible to delete any
segment of *eryA* by ligation of two non-contiguous PCR-generated
fragments and to subsequently construct strains, therefore, devoid of any
or all portions of any module. Such strains deleted of a full module can be

10   employed for reintroduction of either the same or a different module at a
different location. It is possible, therefore, to determine the novel
structures desired and then create a series of *Sac. erythraea* strains
containing the corresponding arrangements of *eryA* modules that would
produce said novel structures that are included within the scope of the

15   present invention. Additional examples of novel compounds produced
from the construction of Type III alterations include but are not limited to
11-deoxyerythromycin, resulting from the insertion of the *eryA* segment
encoding DH4 and ER4 in module 2.

     Moreover, it will also be apparent that two or more modules can be

20   excised and introduced into various sites of the *Sac. erythraea*
chromosome to produce novel polyketides of predicted structure such as
the introduction of the *eryA* segment encoding DH4 and ER4 in both
module 1 and module 2 to result in the production of 14(R)[1-
hydroxypropyl]11-deoxyerythromycin A. All combinations, therefore, of

25   Type III alterations and the strains of *Sac. erythraea* that carry said
alterations as well as the polyketides produced from said strains are
included within the scope of the present invention.

     In addition, it is also possible to create combinations of Type I, Type
II and Type III alterations and insert such alterations into *Sac. erythraea* to

30   produce novel polyketides. Examples of such combinations include but
are not limited to the following. The combination of a Type I alteration,
such as an alteration in DH4 and a Type II alteration, such as a mutation
in the KS1 to result in the formation of (14S,15S)14-[1-hydroxyethyl]-7-
hydroxyerythromycin A when the strain of *Sac. erythraea* carrying such

35   alterations is fed with the compound (2S,3R,4S,5S)3,5-dihydroxy-2,4-
dimethylhexanoic acid ethyl ester. The combination of a Type I alteration,
such as an alteration in DH4 and a Type III alteration, such as found in
Sac. erythraea strain AM412, wherein a copy of the DNA segment of

module 4 is introduced in module 1, such that the *Sac. erythraea* strain so constructed produces the compound 7-hydroxy-14-propylerythromycin A. All combinations of two or more alterations of Type I, Type II and Type III alterations, the *Sac. erythraea* strains that carry such alterations, and the

5 polyketides produced from such strains are included within the scope of the present invention. It will also occur to those skilled in the art that novel structures can be produced by altering the specificity of the acyltransferase functions in any module. Examples include: replacement of the acyltransferase domains of modules 1, 2, 3, 4, 5,or 6 in *eryA* with

10 those of modules 4, 4, 2, 2, 2, and 4, respectively, to result in the production of 12-epierythromycin A,10-epierythromycin A, 8-epierythromycin A, 6-epierythromycin A, 4-epierythromycin A and 2-epierythromycin A, respectively, that are included within the scope of the present invention.

15 It should be emphasized that the introduction of an entire *eryA* module at a different location, as exemplified for the construction of *Sac. erythraea* strains AM412 and AM512 in Examples 29 and 35, respectively, does not rely on homologous recombination between the incoming *eryA* module and the host chromosome. Rather, gene conversion of the host

20 allele with the *eryA* allele residing on the multicopy plasmid requires DNA sequences homologous to the host allele flanking the incoming module. Thus, any module carrying the desired specificities, either from homologous or heterologous sources, can be employed in gene conversion of the host allele, provided that is flanked by segments of

25 homology. It will occur to those skilled in the art, therefore, that, given the large number of natural polyketide molecules existing, a wide variety of additional novel molecules of predicted structure can be produced in Type III mutants containing an additional module of desired specificities or where an endogenous module is replaced by an exogenous one. The

30 length of the acyl chain can be easily controlled by suitably changing the number of modules involved in its synthesis. Similarly, the introduction of keto, hydroxy, enoyl, or methylene groups at specific points along the acyl chain can be easily achieved by introducing the proper b-carbon processing functions (β-ketoreductase, dehydratase and enoylreductase) in

35 the required modules. Exogenous modules constitute the source of specificities for starter and extender units other than those employed by *Sac. erythraea* for erythromycin biosynthesis, making it thereby possible to employ, for example, malonylCoA or (2R)- or (2S)ethylmalonylCoA, etc.

as extender units, and acetyl CoA, butyryl CoA, etc. as the starter unit. The result will be the formation of erythromycin analogs containing the desired functional groups and side chains with the desired stereochemistry. As an extension of the examples reported with *eryA* , the

5   construction of a *Sac. erythraea* strain carrying a heterologous module inserted into *eryA* requires: (i) cloning of the genes from any other Actinomyces producing a polyketide with desired structural features; (ii) mapping of the modular organization of the cloned genes by low stringency hybridization and restriction analysis; (iii) locating the module

10   carrying the desired specificities by partial sequencing; (iv) precise excision of the desired genetic element and cloning into a vector suitable for gene conversion; (v) construction and transformation of a *Sac. erythraea* strain suitable for gene conversion and screening for the novel compound. Any module, or portion thereof, can thus be precisely excised from the genome

15   of a polyketide-producing microorganism and introduced into suitable *Sac. erythraea* strains to create a novel polyketide of predicted structure. Thus, replacement of the acyltransferase segments of modules 1, 2, 3, 4, 5,or 6 in *eryA* with the acyltransferase segment specific for malonyl CoA, such as can be found in the polyketide synthase genes for the synthesis of

20   pikromycin in Streptomyces venezuelae, to result in the synthesis of 12-norerythromycin A, 10-norerythromycin A, 8-norerythromycin A, 6-norerythromycin A, 4-norerythromycin A and 2-norerythromycin A, respectively, that are included within the scope of the present invention. In addition, replacement of the acyltransferase segments of modules 1, 2, 3,

25   4, 5,or 6 in *eryA* with an acyltransferase specific for (2R)-ethylmalonyl CoA, such as can be found in the polyketide synthase genes for the synthesis of spiramycin in Streptomyces ambofasciens, will result in the formation of 12-homoerythromycin A, 10-homoerythromycin A, 8-epihomoerythromycin A, 6-epihomoerythromycin A, 4-

30   epihomoerythromycin A and 2-homoerythromycin A, respectively, all of which are included within the scope of the present invention. Similarly, introduction of acyltransferase segments carrying desired specificities for the starter or extender unit into *eryA* DNA that results in the synthesis of novel compounds are included within the scope of the present invention.

35   The erythromycin analogs produced by the method of this invention are structurally similar to known antibacterial and prokinetic agents.

It will also occur to those skilled in the art that genetic manipulations described herein need not be limited to *Sac. erythraea*.

Suitable hosts are any other polyketide-producing Actinomyces where
DNA can be precisely inserted into the chromosome. Hence, the choice of
a convenient host is based solely on the relatedness of the novel
polyketide to a natural counterpart so as to minimize the number of

5   module rearrangements required for its biosynthesis. Therefore, Type I,
Type II and Type III alterations can be constructed in other Actinomyces
employing either endogenous or exogenous modules to produce novel
polyketides employing strategies analogous to those described herein for
Sac. erythraea. Thus all Type I, Type II or Type III mutations or various

10 combinations thereof constructed in other actinomycetes according to the
principles described herein, and the respective polyketides produced from
such strains, are included within the scope of the present invention.
Examples of polyketides that can be altered by creating Type I, Type II or
Type III changes in the producing microorganisms include, but are not ·

15 limited to macrolide antibiotics such as erythromycin, tylosin, spiramycin,
etc.; ansamacrolides such as rifamycins, maytansines, etc.; polyketide
antibiotics such as tetracycline; polyethers such as monesin, salinomycin,
etc.; polyenes such as candicidin, amphothericins; immunosuppressants
such as FK506, ascomycin, rapamycin, etc. and other complex polyketides

20 such as avermectin.

Whereas the novel derivatives or modifications of erythromycin
described herein have been specified as the A derivatives, such as 7-
hydroxyerythromycin A, 11-oxo-11-deoxyerythromycin A, 14[1(1-
hydroxypropyl)]erythromycin A, etc., those skilled in the art understand

25 that the wild type strain of *Sac. erythraea* produces a family of
erythromycin compounds, including erythromycin A, erythromycin B,
erythromycin C and erythromycin D. Thus, modified strains of *Sac.
erythraea*, such as strain AKR2, for example, would be expected to produce
the corresponding members of the 11-oxo-11-deoxyerythromycin family,

30 including 11-oxo-11-deoxyerythromycin A, 11-oxo-11-deoxyerythromycin
B, 11-oxo-11-deoxyerythromycin C, and 11-oxo-11-deoxyerythromycin D.
Similarly, strain AM412 would be expected to produce not only 14(1-
propyl)erythromycin A but also the other members of the 14(1-
propyl)erythromycin family including 14(1-propyl)erythromycin B, 14(1-

35 propyl)erythromycin C and 14(1-propyl)erythromycin D. Similarly, all
other modified strains of *Sac. erythraea* described herein that produce
novel erythromycin derivatives would be expected to produce the A, B, C,
and D forms of said derivatives. Therefore, all members of the family of

each of the novel polyketides described herein are included within the
scope of the present invention.

Variations and modifications of the methods for obtaining the
desired plasmids, hosts for cloning and choices of vectors and segments of

5   *eryA* DNA to clone and modify, other than those described herein that
result in substantially the same strains and same products as those
described herein will occur to those skilled in the art. For example,
although we have described the use of the plasmids pWH3 and pWHM4
as E. coli-Sac. erythraea shuttle vectors, other vectors can be employed

10 wherein all or part of pWHM3 or pWHM4 is replaced by other DNA
segments that function in a similar manner, such as replacing the pUC19
component of pWHM3 and pWHM4 with pBR322, available from BRL,
employing different segments of the pIJ101 or pJV1 replicons in pWHM3
and pWHM4, respectively, or employing selectable markers other than

15 thiostrepton- and ampicillin-resistance.   These are just few of a long list of
possible examples all of which are included within the scope of the
present invention. Similarly, the segments of the *eryA* locus subcloned
into pWHM3 for generating strains AKS1, AKS2, etc. specified herein can
readily be substituted for other segments of different length encoding the

20 same functions, either produced by PCR-amplification of genomic DNA or
of an isolated clone, or by isolating suitable restriction fragments from *Sac.
erythraea*. In the same way, it is possible to create *eryA* strains carrying
mutations functionally equivalent to those described herein by deleting
different portions of the corresponding genes, by creating insertions into

25 them, or by site-directed mutagenesis of specific nucleotide residues.
Moreover, *Sac. erythraea* strains with mutant alleles other than the β–
ketoacyl ACP synthase portions of *eryA* can be employed as hosts for gene
conversion; Type III mutants can be constructed by double reciprocal
crossover as exemplified for Type I and Type II mutants rather than by the

30 gene conversion method described herein. Additional modifications
include changes in the restriction sites used for cloning or in the general
methodologies described above. All such changes are included in the
scope of the invention. It will also occur to those skilled in the art that
different methods are available to ferment *Sac. erythraea*, to extract the

35 novel polyketides specified herein, and to synthesize substrate analogs,
and that all such methods are also included within the scope of the
present invention.

It will be apparent that many modifications and variations of the invention as set forth herein are possible without departing from the spirit and scope thereof, and that, accordingly, such limitations are imposed only as indicated by the appended claims.

5

39

What is claimed is:

1.   A method for directing the biosynthesis of specific polyketide
analogs by genetic manipulation of a polyketide-producing
5   microorganism, said method comprising the steps of:
       (1) isolating a polyketide biosynthetic gene-containing DNA
sequence;
       (2) identifying enzymatic activities associated within said gene-
containing DNA sequence;
10   (3) introducing one or more specified changes into said gene-
containing DNA sequence which codes for one of said enzymatic activities
resulting in an altered DNA sequence;
       (4) introducing said altered DNA sequence into a polyketide-
producing microorganism to replace the original sequence;
15   (5) growing a culture of the altered microorganism under
conditions suitable for the formation of the specific polyketide analog; and
       (6) isolating said specific polyketide analog from the culture.

2.   The method of claim 1 wherein said polyketide biosynthetic gene-
20   containing DNA sequence comprises genes which encode the enzymatic
activities comprising a polyketide synthase.

3.   The method of claim 2, wherein said polyketide synthase enzymatic
activities comprise β-ketoreductase, dehydratase, acyl carrier protein,
25   enoylreductase, β-ketoacyl ACP synthase, and acyltransferase.

4.   The method of claim 1 wherein said alteration which occurs in the
DNA sequence results in the inactivation of one or more enzymatic
activities involved in the processing of the β-carbonyl of said polyketide.
30

5.   The method of claim 4, wherein said inactivated enzymatic
activities affecting the processing of the β-carbonyl of said polyketide
comprise β-ketoreductase, dehydratase, and enoylreductase.

35   6.   The method of claim 4 wherein said alteration in the DNA
sequence results in the addition of one or more enzymatic activities
involved in the β-carbonyl processing of said polyketide.

7.   The method of claim 6 wherein said additional enzymatic activities
are selected from the group consisting of β-ketoreductase, β-ketoreductase
and dehydratase, and β-ketoreductase, dehydratase and enoylreductase.

8.   The method of claim 1 wherein said alteration occurring in the
DNA segment results in the inactivation of one or more enzymatic
activities involved in the condensation of carbon units to the nascent
polyketide structure.

9.   The method of claim 8 wherein said enzymatic activities affecting
the condensation of carbon units to the nascent polyketide structure
comprise β-ketoacyl ACP synthase, acyl carrier protein and acyltransferase.

10.   The method of claim 1 wherein said alteration in the DNA
sequence results in the change of the length of the polyketide synthesized.

11.   The method of claim 10 wherein said alteration results in the
increase of the length of the polyketide.

12.   The method of claim 11 wherein said alteration comprises the
addition of DNA sequences encoding the enzymatic activities consisting of
acyltransferase, acyl carrier protein and β-ketoacyl ACP synthase.

13.   The method of claim 10 wherein said alteration results in the
decrease of the length of the polyketide.

14.   The method of claim 13 wherein said alteration consists of the
deletion of a DNA segment between two sequences encoding
corresponding enzymatic activities.

15.   The method of claim 14 wherein said corresponding enzymatic
activities are selected from the group consisting of β-ketoreductases,
dehydratases, acyl carrier proteins, enoylreductases, β-ketoacyl ACP
synthases, and acyltransferases.

16.   The method of claim 1 wherein said alteration consists in the
replacement of the DNA segment encoding an acyltransferase with a DNA
segment encoding an acyltransferase of different specificity.

17.   The method of claim 1 wherein said DNA sequence is isolated from a species from the *Actinomycetales* family.

5   18.   The method of claim 17 wherein said DNA sequence is isolated from a genus selected from the group consisting of *Actinomyces, Dactylosporangium, Micromonospora, Nocardia, Sac., Streptoverticillium,* and *Streptomyces.*

10   19.   The method of claim 17 wherein said genus is selected from the group consisting of *Saccharapolyspora* and *Streptomyces.*

20.   The method of claim 19 wherein said genus is *Saccharapolyspora* and the species is *erythraea.*

15

21.   The method of claim 19 wherein said genus is *Streptomyces* and the species is *hygroscopicus.*

22.   The method of claim 1 wherein said polyketide is selected from the
20   group consisting of macrolides, tetracyclines, polyethers, polyenes, ansaymcins and derivatives or analogs thereof.

23.   The method of claim 22 wherein said polyketide is a macrolide.

25   24.   The method of claim 23 wherein said macrolide is an erythromycin.

25.   The method of claim 24 wherein said erythromycin analog is selected from the group consisting of 11-oxo-11-deoxyerythromycin A, 7-hydroxyerythromycin A, 6-deoxy-7-hydroxyerythromycin A, 7-
30   oxoerythromycin A, 3-oxo-3-deoxy-5-desosaminylerythronolide A, Δ-6,7-. anhydroerythromycin A, ((14S,15S)14(1-hydroxyethyl)erythromycin A, 11-epifluoro-15-norerythromycinA, 14-(1-propyl)erythromycin A, 14(1-propyl)erythromycin A, and 14[1(1-hydroxypropyl)]erythromycin A.

35   26.   The method of claim 1 wherein said DNA sequence, designated *eryA*, encodes the enzymatic activities associated with the formation of 6-deoxyerythronolide B.

27.   The method of claim 26 wherein said DNA sequence comprises:
the DNA sequence of Figure 2.

28.   The method of claim 1 wherein said gene-containing DNA
5   sequence encodes one or more enzymatic activities in the rapamycin
biosynthetic pathway.

29.   The method of claim 23 wherein said macrolide is a rapamycin
analog.

10

30.   A compound selected from the group consisting of 7-
hydroxyerythromycin A; 6-deoxy-7-hydroxyerythromycin A; 7-
oxoerythromycin A, 3-oxo-3-deoxy-5-desosaminyl-erythronolide A; Δ-6,7-
anhydroerythromycin A; ((14S,15S)14(1-hydroxyethyl)erythromycin A; 11-
15   epifluoro-15-norerythromycin A; 14-(1-propyl)erythromycin A; 14(1-
propyl)erythromycin A; and 14[1(1-hydroxypropyl)]erythromycin A.

●

FIG.1

FIG.2-1
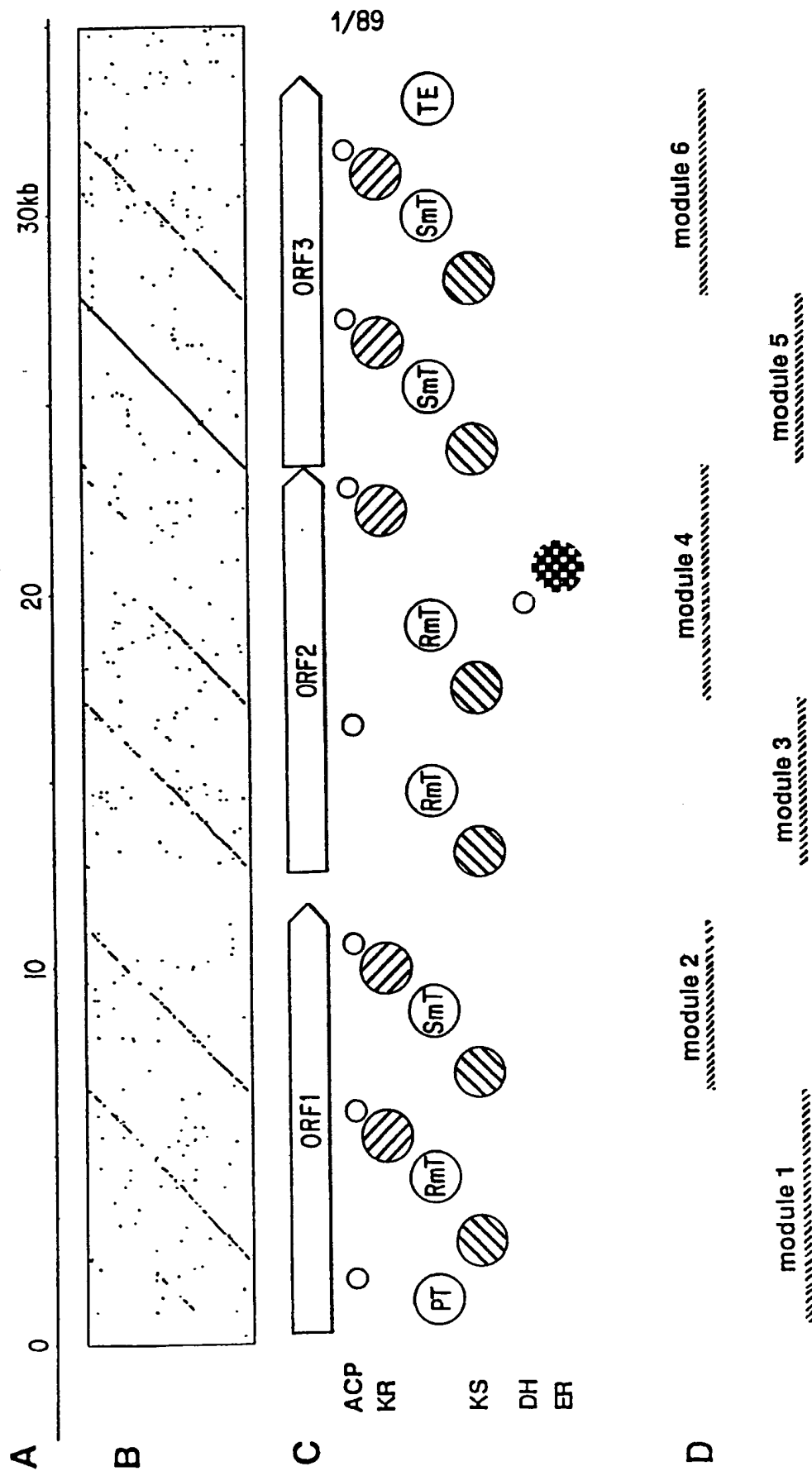
```
      L   L   G   E   S   R   V   F   A   A   A   M   D   A   C   A   R   A   F   E   -
1021  AGCCCGTGACCGACTGGACGCTGGCGGCCAGTCGGCGCCGGCG
      --------+---------+---------+---------+ 1080

      P   V   T   D   W   T   L   A   Q   V   L   D   S   P   E   Q   S   R   R   V   -
1081  TCGAGGTCGTCCAGCCCGCCCTGTTCGCGGTGCAGACGTCGCTGGCCGCGCTCTGGCGCT
      --------+---------+---------+---------+ 1140

      E   V   V   Q   P   A   L   F   A   V   Q   T   S   L   A   A   L   W   R   S   -
1141  CCTTCGGCGTGACCCCGACGCCGTGGTGGGCCACAGCATCGGCGAGCTGGCCGCCGCGC
      --------+---------+---------+---------+ 1200

      F   G   V   T   P   D   A   V   V   G   H   S   I   G   E   L   A   A   A   H   -
1201  ACGTGTGCGGTGCGGCCGCGGTGCCGCCGACGCCGCGGCGCCGCTGTGTGGAGCCGCG
      --------+---------+---------+---------+ 1260

      V   C   G   A   A   G   A   A   D   A   A   R   A   A   A   L   W   S   R   E   -
1261  AGATGATTCCGTTGGGTGGGCAACGGGCGACATGGCAGCCGTCGCCGTCTCCGCCGACGAGA
      --------+---------+---------+---------+ 1320

      M   I   P   L   V   G   N   G   D   M   A   A   V   A   L   S   A   D   E   I   -
1321  TCGAGCCGCGCGCATCGCCCGCCGGTGGGACGACGACGTGGTGCTGCCGGGGTCAACGGTCCGC
      --------+---------+---------+---------+ 1380
```

FIG.2-2

```
          E   P   R   I   A   R   W   D   D   D   V   V   L   A   G   V   N   G   P   R   -
1381   GCTCGGTTCTGCTGACCGGGTCGCCGGAACCGGTCGCCGCGGGTCCAGGAGCTCTCGG
       ---------+---------+---------+---------+---------+---------+   1440

          S   V   L   L   T   G   S   P   E   P   V   A   R   R   V   Q   E   L   S   A   -
1441   CCGAGGGGGTCCGCGCACAGGTCATCAATGTGTCGATGGCGGCACTCGGCGCAGGTCG
       ---------+---------+---------+---------+---------+---------+   1500

          E   G   V   R   A   Q   V   I   N   V   S   M   A   A   H   S   A   Q   V   D   -
1501   ACGACATCGCCGAGGGGGATGCGCTCGGCCCTGGTGGTTCGCGCCCGGTGGCTCGGAGG
       ---------+---------+---------+---------+---------+---------+   1560

          D   I   A   E   G   M   R   S   A   L   A   W   F   A   P   G   G   S   E   V   -
1561   TGCCCCTTCTACGCCAGCCTCACCGGAGGTGCGGTCGGACACGCGGGGAGCTGGTGGCCGACT
       ---------+---------+---------+---------+---------+---------+   1620

          P   F   Y   A   S   L   T   G   G   A   V   D   T   R   E   L   V   A   D   Y   -
1621   ACTGGCGCCGCAGCTTCCGGCTGCCGGTGCGGCTTCGACGAGGCGGATCCGCGGTCCGGCCCCTGG
       ---------+---------+---------+---------+---------+---------+   1680

          W   R   S   F   R   L   P   V   R   F   D   E   A   I   R   S   A   L   E   -
```

```
1681  AGGTCGGTCCCGGCACGTTCGTCGAAGCGGAGCCCGGCACCCGGTGCTGGCCGCGCTCC  1740
      ---------+---------+---------+---------+---------+---------+
       V  G  P  G  T  F  V  E  A  S  P  H  P  V  L  A  A  A  L  Q   -

1741  AGCAGACGCTCGACGCCGAGGGCTCCTCGGCCGCGGTGGTCCCGACGCTGCAACGCGGGC  1800
      ---------+---------+---------+---------+---------+---------+
       Q  T  L  D  A  E  G  S  S  A  A  V  V  P  T  L  Q  R  G  Q   -

1801  AGGGGCGGCATGCGGCGGTTCCTGCTGGCCGCGGCCCAGGCCGTTCACCGGCGGTGGCCG  1860
      ---------+---------+---------+---------+---------+---------+
       G  G  M  R  R  F  L  A  A  A  Q  A  F  T  G  G  A  V   -

1861  TCGACTGGACCGCCGCCTACGACGACGTGGGGCCGAACCCGGCTCTCTGCCGAGTTCGC  1920
      ---------+---------+---------+---------+---------+---------+
       D  W  T  A  A  Y  D  D  V  G  P  N  P  A  L  C  R  S  S  R   -

1921  GCCGGCCGAGGAGGAAGAGACGAGCCCGAGTCCGGCGTCGACTGGAACGCGCCACCGCA  1980
      ---------+---------+---------+---------+---------+---------+
       R  P  R  R  K  T  S  R  P  S  P  A  S  T  G  T  R  H  R  T   -

1981  CGTGCTGCGAGCGGCTGCTCGCGGGTGTCGTCAACGGCGAGACCGCCGGTTGGGGCCGCG  2040
      ---------+---------+---------+---------+---------+---------+
       C  C  E  R  L  L  A  V  V  N  G  E  T  A  A  L  A  G  R  E   -
```

FIG.2-3

FIG.2-4

```
2041  AAGCCGACGCCGAGGCCACGTTCCGCGAGCTGGGCCTGGACTCGGTGCTGGCCGCGCAGC
      ----+----+----+----+----+----+  2100
         A  D  A  E  A  T  F  R  E  L  G  L  D  S  V  L  A  A  Q  L -

2101  TGGCGGCCAAGGTGAGCGCCGCGATCGGGCGCGAGGTCAACATCGCCCTGCTCTACGACC
      ----+----+----+----+----+----+  2160
         R  A  K  V  S  A  A  I  G  R  E  V  N  I  A  L  L  Y  D  H -

2161  ACCCGACTCCGCGTGCGCTCGCGGAAGCACTCGCGGGGGAACCGAGGTCGCCACAACGGG
      ----+----+----+----+----+----+  2220
         P  T  P  R  A  L  A  E  A  L  A  A  G  T  E  V  A  Q  R  E -

2221  AAACCCGGCGCGGACCAACGAAGCGGGCCCGGCGTCGCGGTCGCGGATGG
      ----+----+----+----+----+----+  2280
         T  R  A  R  T  N  E  A  A  P  G  E  P  V  A  V  V  A  M  A -

2281  CCTGCCGGCTGCCCGGCGGTGTGAGCACCCCGGAGGAGTTCTGGGAGCTGCTGTCGGGAGG
      ----+----+----+----+----+----+  2340
         C  R  L  P  G  G  V  S  T  P  E  E  F  W  E  L  L  S  E  G -

2341  GCCGCGACGCGGTCGCGGGACTGCCGACCGACCGGACTCCGCTGTTCC
      ----+----+----+----+----+----+  2400
```

```
        R  D  A  V  A  G  L  P  T  D  R  G  W  D  L  D  S  L  F  H  -
        ACCCCGACCCCACGCGCTCGGGCCACCGGCCACCAGCGCGGCGGGGCGGTTCCTGACCGAGG
   2401 ---------+---------+---------+---------+---------+---------+ 2460

        P  D  P  T  R  S  G  T  A  H  Q  R  G  G  G  G  F  L  T  E  A  -
        CGACCGGCGTTCGACCCGGCCCTTCTTCGGCCATGTCCCCGCGGAGGGCCTGGCCGTCGACC
   2461 ---------+---------+---------+---------+---------+---------+ 2520

        T  A  F  D  P  A  F  F  G  M  S  P  R  E  A  L  A  V  D  P  -
        CGCAGCAGCGGCTCATGCTCGAGCTCTCCTGGGAAGTGCTGGAACGGGCGGGAATCCCGC
   2521 ---------+---------+---------+---------+---------+---------+ 2580

        Q  Q  R  L  M  L  E  L  S  W  E  V  L  E  R  A  G  I  P  P  -
        CGACCTCGTTGCAGGCCTCGCCCACTGGCGTGTTCGTCGGCCTGATCCCGCAGGAGTACG
   2581 ---------+---------+---------+---------+---------+---------+ 2640

        T  S  L  Q  A  S  P  T  G  V  F  V  G  L  I  P  Q  E  Y  G  -
        GCCCGCGGCTGGCCGAGGGCGGCGGCGTCGAGGGCGTCGAGGGCTACCTGATGACCGGTACGACCA
   2641 ---------+---------+---------+---------+---------+---------+ 2700

        P  R  L  A  E  G  G  E  G  V  E  G  Y  L  M  T  G  T  T  T  -
        CGAGCGTCGCCTTCCGGCCCATCGCCTACACGCTCGGAGGCCCGGCGATCAGCG
   2701 ---------+---------+---------+---------+---------+---------+ 2760
```

FIG.2-5

FIG.2-6

```
           S   V   A   S   G   R   I   A   Y   T   L   G   L   E   G   P   A   I   S   V  -
2761  TGGACACCGGCCGTGCTCGTCCTCGCTGGTCGGGTGCACCTGGGCGGTGCCAGTCGCTGCGGC
      ---------+---------+---------+---------+---------+---------+  2820

           D   T   A   C   S   S   S   L   V   A   V   H   L   A   C   Q   S   L   R   R  -
2821  GCGGCGGAGTCGTCGCTGGCGATGGCAGGCGGGTGTCACGGTGTGCACGGTGATGCCGACGCCCGGCATGC
      ---------+---------+---------+---------+---------+---------+  2880

           G   E   S   S   L   A   M   A   G   G   V   T   V   M   P   T   P   G   M   L  -
2881  TGGTGGACTTCAGCCGGATGAACTCGCTGGCCCGGACGGCCGGTGCAAGGCTTTCTCCG
      ---------+---------+---------+---------+---------+---------+  2940

           V   D   F   S   R   M   N   S   L   A   P   D   G   R   C   K   A   F   S   A  -
2941  CCGGCGCCAACGGTTTCGGCATGGCCGAGGGCGCCGGGATGCTCCTGCTGGAGCGGCTTT
      ---------+---------+---------+---------+---------+---------+  3000

           G   A   N   G   F   G   M   A   E   G   A   G   M   L   L   E   R   L   S  -
3001  CGGACGCCCGCCGCAACGGCCACCCGGTGCTCGCCGTGCTCAGGGGACGGCGGTCAACT
      ---------+---------+---------+---------+---------+---------+  3060

           D   A   R   R   N   G   H   P   V   L   A   V   L   R   G   T   A   V   N   S  -
      CCGACGGCGGAGCAACGGGCTGTCGGGCGGCCCAACGGGCGGGCCAGGTGCGGGTCATCC
```

FIG.2−7

```
3061 ----+----+----+----+----+----+ 3120
      D  G  A  S  N  G  L  S  A  P  N  G  R  A  Q  V  R  V  I  Q  -
     AGCAGGCGCTGGCAGAGTCCGGTCTCGGGCCCGCCGACATCGACGCCGTCGAGGCGCACG
3121 ----+----+----+----+----+----+ 3180
      Q  A  L  A  E  S  G  L  G  P  A  D  I  D  A  V  E  A  H  G  -
     GCACCGGTACCCGACTCGGCGACCCGATCGAGGCGCGCTGTTCGAGGCGTACGGGGC
3181 ----+----+----+----+----+----+ 3240
      T  G  T  R  L  G  D  P  I  E  A  R  A  L  F  E  A  Y  G  R  -
     GCGACCGGCGAGCAGCCGCTGCACCTGGGCTCGGTCAAGTCCAACCTCGGCCACCCAGG
3241 ----+----+----+----+----+----+ 3300
      D  R  E  Q  P  L  H  L  G  S  V  K  S  N  L  G  H  T  Q  A  -
     CGGCCGCCGGGTGTTGCCGGGGTGATCAAGATGGTGCTGGCGCGCGGGGCCACCCTTC
3301 ----+----+----+----+----+----+ 3360
      A  A  G  V  A  G  V  I  K  M  V  L  A  M  R  A  G  T  L  P  -
     CCCGCCACTCTGCACGCATCGGAGCGGCGGTCGAAGGAGATCGACTGGTCATCCGGTGCGGATCA
3361 ----+----+----+----+----+----+ 3420
      R  T  L  H  A  S  E  R  S  K  E  I  D  W  S  S  G  A  I  S  -
```

FIG.2-8

```
3421 GCCTGCTCGACGAGCCGGAGCCGTGGCCGGCGCCGCGCCGACCGCGCCGGGGCGGGGGTCT 3480
     ----+----+----+----+----+----+----+----+----+----+----+----+
      L  D  E  P  E  P  W  P  A  G  A  R  P  R  R  A  G  V  S  -

3481 CGTCGTTCGGCATCAGCGGCACGAACGGCCATCATCGAGGAAGCTCCGCAGGTCG 3540
     ----+----+----+----+----+----+----+----+----+----+----+----+
      S  F  G  I  S  G  T  N  A  H  I  E  E  A  P  Q  V  V  -

3541 TCGAAGGCGAGCGGGTCGAGGCCGGACGTCGTGGCCCTGGGTGCTTTCGGCGAGCA 3600
     ----+----+----+----+----+----+----+----+----+----+----+----+
      E  G  E  R  V  E  A  G  D  V  V  A  P  W  V  L  S  A  S  S  -

3601 GCGCGGAAGGTCTGCGCCCAGGGCGGCCACCTGCGCGCGAGCACCCCG 3660
     ----+----+----+----+----+----+----+----+----+----+----+----+
      A  E  G  L  R  A  Q  A  A  R  L  A  A  H  L  R  E  H  P  G  -

3661 GTCAGGACCCGCGCGACATCGCGCTACTCGCTCCGCGACGGGACGGCCGCTGCCCCACC 3720
     ----+----+----+----+----+----+----+----+----+----+----+----+
      Q  D  P  R  D  I  A  Y  S  L  A  T  G  R  A  A  L  P  H  R  -

3721 GCGCCGGCCTTCGCCCCCGTCGACGAGTCCGCGGCGTCTGCTCGACGGTCTCGCGA 3780
     ----+----+----+----+----+----+----+----+----+----+----+----+
      A  A  F  A  P  V  D  E  S  A  A  L  R  V  L  D  G  L  A  T  -
```

```
3781 CGGGAAACGCCGACGGTGCCGCCGTTGGAACGAGCCGGGGCCAGCAGCGCCGTCTTCG 3840
     ----+----+----+----+----+----+----+----+----+----+----+
       G  N  A  D  G  A  A  V  G  T  S  R  A  Q  Q  R  A  V  F  V  -

3841 TCTTCCCGGGGCAGGGTTGGCAGTGGGCGGGCATGGCCGTCGACCTGCTCGACACCTCCC 3900
     ----+----+----+----+----+----+----+----+----+----+----+
       F  P  G  Q  G  W  Q  W  A  G  M  A  V  D  L  L  D  T  S  P  -

3901 CGGTTTTCGCAGCCGCCGTTGCGCGGCCGAGTGCGCCTCGAACCGCCATCTCTGGACTTCG 3960
     ----+----+----+----+----+----+----+----+----+----+----+
       V  F  A  A  A  L  R  E  C  A  D  A  L  E  P  H  L  D  F  E  -

3961 AGGTGATCCCGTTCCTGCGCGCGGAAGCCGCGGGAGCCGGGAGGCAGGACGCGGGCGTCGA 4020
     ----+----+----+----+----+----+----+----+----+----+----+
       V  I  P  F  L  R  A  E  A  A  R  R  E  Q  D  A  A  L  S  T  -

4021 CCGAGCGGCGTGGACGTGGTGCAGCCCGTGATGTTCGCGGTCATGGTCTCGCTGGCGTCGA 4080
     ----+----+----+----+----+----+----+----+----+----+----+
       E  R  V  D  V  V  Q  P  V  M  F  A  V  M  V  S  L  A  S  M  -

4081 TGTGGCGAGCCCACGGCGTCGAGCCGGCCGTCATCGGCCACTCCCAGGGCCGAGATCG 4140
     ----+----+----+----+----+----+----+----+----+----+----+
       W  R  A  H  G  V  E  P  A  A  V  I  G  H  S  Q  G  E  I  A  -
```

FIG.2-9

```
4141 CCGCCGGCGTGGCGTCGGGGGGCGGCGCTCTCGCTGGACGACGCGCGGCGCGTGGTCGCGCTGC 4200
     ---------+---------+---------+---------+---------+---------+
      A  A  C  V  A  G  A  L  S  L  D  D  A  A  R  V  V  A  L  R  -

4201 GCAGCCGCGTCATCGCCACCATGCCCGGGAACAAGGGCATGGCCTCGATCGCCGCTCCGG 4260
     ---------+---------+---------+---------+---------+---------+
      S  R  V  I  A  T  M  P  G  N  K  G  M  A  S  I  A  A  P  A  -

4261 CCGGGCGAAGTCCGGCGGCGAATCGGTGACCGGTCGCCGCGTCAACGGTCCGC 4320
     ---------+---------+---------+---------+---------+---------+
      G  E  V  R  A  R  I  G  D  R  V  E  I  A  A  V  N  G  P  R  -

4321 GCTCGGTGGTGGTCGCGGGCGACAGCGACGAACTGGACCGGCTGGTCGCTTCCTGCACCA 4380
     ---------+---------+---------+---------+---------+---------+
      S  V  V  V  A  G  D  S  D  E  L  D  R  L  V  A  S  C  T  T  -

4381 CCGAGTGCATCCGCGCCAAGCGGCTGGCCGTGGACTACGCGTCGCACTCCTCGCACGTCG 4440
     ---------+---------+---------+---------+---------+---------+
      E  C  I  R  A  K  R  L  A  V  D  Y  A  S  H  S  S  H  V  E  -

4441 AGACGGATCCGAGAGACGCACTGCACGCCGAGCTGGGAGAGGACTTCCACCCGCTGCCGGGGGT 4500
     ---------+---------+---------+---------+---------+---------+
```

FIG.2-10

FIG.2-11

```
        T  I  R  D  A  L  H  A  E  L  G  E  D  F  H  P  L  P  G  F  -
        TCGTGCCCTTCTTCTCCACCGTCACCGGGGCGCTGGACGGCAGCCGGACGGAGCTCGACGCCG
4501    ------+---------+---------+---------+---------+---------+   4560

        V  P  F  S  T  V  T  G  R  W  T  Q  P  D  E  L  D  A  G  -
        GGTACTGGTACCGGAACCTGCGCGCCGCCGGGGTTCGGCGACGCCGTCCGTGCGCTCG
4561    ------+---------+---------+---------+---------+---------+   4620

        Y  W  Y  R  N  L  R  R  T  V  R  F  A  D  A  V  R  A  L  A  -
        CCGAGCAGGGATATCGCACGTTCCTGGAGGTCAGCGCCCACCCGATCCTCACCGCCGCGA
4621    ------+---------+---------+---------+---------+---------+   4680

        E  Q  G  Y  R  T  F  F  L  E  V  S  A  H  P  I  L  T  A  A  I  -
        TCGAGGAGATCGGGCGACGGATCGGGCGCCGCGACCTCTCCGCCATCCATTCGCTGCGCCGCG
4681    ------+---------+---------+---------+---------+---------+   4740

        E  E  I  G  D  G  S  G  A  D  L  S  A  I  H  S  L  R  R  G  -
        GTGACGGGCAGCCTCGCGACTTCGGCGAAGCGCTCTCCCGGCGTTCGCGCCGGTGTCG
4741    ------+---------+---------+---------+---------+---------+   4800

        D  G  S  L  A  D  F  G  E  A  L  S  R  A  F  A  A  G  V  A  -
        CGGTGGACTGGGAGTCGGTGCACCTGGGCACGCGGAGCACCGGGTGCCCTTGCCCACCT
4801    ------+---------+---------+---------+---------+---------+   4860
```

FIG.2-12

```
         V  D  W  E  S  V  H  L  G  T  G  A  R  R  V  P  L  P  T  Y  -
4861  ACCCGTTCCAGCGGCGAGCGCGGTCTGGCTCGAACCGAAGCCGGTGGGCGGCCGGTCCACCG
      ----+---------+---------+---------+---------+---------+---------+  4920

         P  F  Q  R  E  R  V  W  L  E  P  K  P  V  A  R  R  S  T  E  -
4921  AGGTCGACGAGGTTCCGCGCTGCGCTACCGCATCGAGTGGGCGCCCACCGGTGCCGGTG
      ----+---------+---------+---------+---------+---------+---------+  4980

         V  D  E  V  S  A  L  R  Y  R  I  E  W  R  P  T  G  A  G  E  -
4981  AACCCGCCCGGCTCGACGGCCACCTGGCTGGTGGCGAAGTACGCCGGACCGGAACCGGACGAGA
      ----+---------+---------+---------+---------+---------+---------+  5040

         P  A  R  L  D  G  T  W  L  V  A  K  Y  A  G  T  A  D  E  T  -
5041  CGAGCACCGCGGCTCGGGAGCCCTGGGAGTCGGCGCGGTCCGCGAACTGGTCG
      ----+---------+---------+---------+---------+---------+---------+  5100

         S  T  A  R  E  A  L  E  S  A  G  A  R  V  R  E  L  V  V  -
5101  TGGACGCCCGCTGCGGTCGGCGACGAACTCGCCGGAGCGGCTTCGTTCGGTCGGGCGAGGTGG
      ----+---------+---------+---------+---------+---------+---------+  5160

         D  A  R  C  G  R  D  E  L  A  E  R  L  R  S  V  G  E  V  A  -
```

```
5161  CAGGAGTGCTGTCCCTGCTCGCGGGTGGACGAAGCCGGAGGAGGCCGCTCGCGC  5220
      ------+---------+---------+---------+---------+---------+
       G  V  L  S  L  L  A  V  D  E  A  E  P  E  E  A  P  L  A  L  -

5221  TGGCTTCGCTGGCGGACACGCTCAGCGCCTCGTGCAGGCGATGGTGTCGGCCGAACTCGGAT  5280
      ------+---------+---------+---------+---------+---------+
       A  S  L  A  D  T  L  S  L  V  Q  A  M  V  S  A  E  L  G  C  -

5281  GTCCGCTGTGTGGACGGTGAAAGCGGCCGTCGCGACGGGCCGTTCGAACGCGTCCGCA  5340
      ------+---------+---------+---------+---------+---------+
       P  L  W  T  V  T  E  S  A  V  A  T  G  P  F  E  R  V  R  N  -

5341  ACGCCGCCCACGGCGCCCTGTGGGGCGTCGGGGGTCATCGCGCTGGAGAACCCCGCCG  5400
      ------+---------+---------+---------+---------+---------+
       A  A  H  G  A  L  W  G  V  G  R  V  I  A  L  E  N  P  A  V  -

5401  TGTGGGGCGGCCTGGTCGACGTGCCCGGCGTCGGCCGAGCTGGCCCGGCACCTCG  5460
      ------+---------+---------+---------+---------+---------+
       W  G  G  L  V  D  V  P  A  G  S  V  A  E  L  A  R  H  L  A  -

5461  CGGCGGGTCGTGTCCGGCGGCGGTGAGGACCAGCTCGCGCTGCGCGCCGACGGGGGTGT  5520
      ------+---------+---------+---------+---------+---------+
       A  V  V  S  G  G  A  G  E  D  Q  L  A  L  R  A  D  G  V  Y  -
```

FIG.2-13

```
5521 ACGGACGCCGGTGGGTGCGCGGCCCCGGCGACCGATGACGAGTGGAAACCCACCG   5580
     ----+----+----+----+----+----+----+----+----+----+----+
       G  R  R  W  V  R  A  A  A  P  A  T  D  D  E  W  K  P  T  G  -

5581 GAACCGGTGCTGGTCACCGGTGGCACGGGCGGTGTCGGCGGCCAGATCGCGCGCTGGCTCG   5640
     ----+----+----+----+----+----+----+----+----+----+----+
       T  V  L  V  T  G  G  T  G  G  V  G  G  Q  I  A  R  W  L  A  -

5641 CCCGGGCGGGGCGCCCCCACCTGCTGCTGGTGAGCCGCAGCGGCCCGGACGCGGACGGCG   5700
     ----+----+----+----+----+----+----+----+----+----+----+
       R  R  G  A  P  H  L  L  L  V  S  R  S  G  P  D  A  D  G  A  -

5701 CCGGGCGAACTGGTCGCCGAGCTCGGAGGCGCTGGGCGCCCCGGACCGTCGCGGCCTGCG   5760
     ----+----+----+----+----+----+----+----+----+----+----+
       G  E  L  V  A  E  L  E  A  L  G  A  R  T  T  V  A  A  C  D  -

5761 ACGTGACCGGACCGGAGTCGGTTCGCGGCATCGGTGACGACGTCCCGC   5820
     ----+----+----+----+----+----+----+----+----+----+----+
       V  T  D  R  E  S  V  R  E  L  L  G  G  I  G  D  D  V  P  L  -

5821 TCTCGGCGGTGTTCCACGCCGCCGCGCTCGACGGCACCGTGGACACCCTCACCG   5880
     ----+----+----+----+----+----+----+----+----+----+----+
```

FIG.2-14

```
          S   A   V   F   H   A   A   A   T   L   D   D   D   G   T   V   D   T   L   T   G   -
       GCGAGCGGCATCGAGAGCGGGGGCAAGTGTCGCGCGCCAAGGTGCTCGGCGGCGCCGCAACCTGCACGAGC
       ----+----+----+----+----+----+----+----+----+----+----+----+          5940
5881   GCGAGCGGCATCGAGAGCGGGGGCAAGTGTCGCGCGCCAAGGTGCTCGGCGGCGCCGCAACCTGCACGAGC

          E   R   I   E   R   A   S   R   A   K   V   L   G   A   R   N   L   H   E   L   -
       TGACGCGGCGAGCTGGACCTGACCGGCCTTCGTCGTGCTGTTCTCGTCCTTCGGCCTTCG
       ----+----+----+----+----+----+----+----+----+----+----+----+          6000
5941   TGACGCGGCGAGCTGGACCTGACCGGCCTTCGTCGTGCTGTTCTCGTCCTTCGGCCTTCG

          T   R   E   L   D   L   T   A   F   V   L   F   S   S   F   A   S   A   F   G   -
       GCGCCCCGGGCTCGGCGGCTACGCGCCCGGCAACGCCTACCTCGACGGCCTCGCCCAGC
       ----+----+----+----+----+----+----+----+----+----+----+----+          6060
6001   GCGCCCCGGGCTCGGCGGCTACGCGCCCGGCAACGCCTACCTCGACGGCCTCGCCCAGC

          A   P   G   L   G   G   Y   A   P   G   N   A   Y   L   D   G   L   A   Q   Q   -
       AGCGGCGGAGCGGACGGACTCCCGCGACCCCGCGTGGGGGCGACGTGGGCGGCAGCG
       ----+----+----+----+----+----+----+----+----+----+----+----+          6120
6061   AGCGGCGGAGCGGACGGACTCCCGCGACCCCGCGTGGGGGCGACGTGGGCGGCAGCG

          R   R   S   D   G   L   P   A   T   A   V   A   W   G   T   W   A   G   S   G   -
       GGATGGCCGGAAGGCGGGTGGCCGACCGCTTCCGCAGCACGGCGTCATCGAGATGCCTC
       ----+----+----+----+----+----+----+----+----+----+----+----+          6180
6121   GGATGGCCGGAAGGCGGGTGGCCGACCGCTTCCGCAGCACGGCGTCATCGAGATGCCTC

          M   A   E   G   A   V   A   D   R   F   R   R   H   G   V   I   E   M   P   P   -
       CCGAGACGGCCTGCCGGGCGTTGCAGAACGCGCTGGACCGGCTCTGCCCGATCG
       ----+----+----+----+----+----+----+----+----+----+----+----+          6240
6181   CCGAGACGGCCTGCCGGGCGTTGCAGAACGCGCTGGACCGGCTCTGCCCGATCG
```

FIG.2—15

FIG.2-16

```
          E  T  A  C  R  A  L  Q  N  A  L  D  R  A  E  V  C  P  I  V  -
      TCATCGACGTCAGGTGGGACCGGTTCCTGCTGCTCGCCTACACCGGCGCAGCGCCCGACCAGGC
6241  ----+----+----+----+----+----+----+----+----+----+----+----+ 6300

          I  D  V  R  W  D  R  F  L  L  A  Y  T  A  Q  R  P  T  R  L  -
      TCTTCGACGAGATCGACGACGCGCGGGCGGGGGCTGCGCCAGGCGCCGGGCCGGAACCGCGGG
6301  ----+----+----+----+----+----+----+----+----+----+----+----+ 6360

          F  D  E  I  D  D  A  R  R  A  A  P  Q  A  P  A  E  P  R  V  -
      TGGGCGGCGCTGGCGTCGCTGCCCGCGCCGGAGAGCGCCGAGGAAGGCGCTGTTCGAGCTCGTGC
6361  ----+----+----+----+----+----+----+----+----+----+----+----+ 6420

          G  A  L  A  S  L  P  A  P  E  R  E  E  A  L  F  E  L  V  R  -
      GCTCGCACGCGGCCGCGCCGTCCTCGGCCACGCGCCTCGGCCGAGCGGGTGCCCGAGCCGACCAGG
6421  ----+----+----+----+----+----+----+----+----+----+----+----+ 6480

          S  H  A  A  A  V  L  G  H  A  S  A  E  R  V  P  A  D  Q  A  -
      CCTTCGCGGAACTCGGCGGTCGACTCGCTGCGGCGCTTGAGCTGCGCAACCGGCTCGGGCG
6481  ----+----+----+----+----+----+----+----+----+----+----+----+ 6540

          F  A  E  L  G  V  D  S  L  S  A  L  E  L  R  N  R  L  G  A  -
      CCGCGACCGGTGTCCGCGCCTGCCGACGAGACGTCTTCGACCACCCGACGTGCGGGACGC
```

FIG.2-17

```
6541 --------+---------+---------+---------+---------+---------+ 6600
          A   T   G   V   R   L   P   T   T   V   F   D   H   P   D   V   R   T   L   -
     TGGCGGCGCACCTGGCCGCCGAACTCGGCGGTGCCGACCGGAGCCGAGCAGGCGCACCGG

6601 --------+---------+---------+---------+---------+---------+ 6660
          A   A   H   L   A   A   E   L   G   G   A   T   G   A   E   Q   A   A   P   A   -
     CGACCACGGCCCCGTCGACGAGCCGATCGCGTCGGCATGGCCGTGCCGGCTGCCCG

6661 --------+---------+---------+---------+---------+---------+ 6720
          T   T   A   P   V   D   E   P   I   A   I   V   G   M   A   C   R   L   P   G   -
     GGGAGGTCGACTCCCCCGGAGCGCTGTGTGGGAGCTGATCACCTCGGGACGCGACTCCGCGG

6721 --------+---------+---------+---------+---------+---------+ 6780
          E   V   D   S   P   E   R   L   W   E   L   I   T   S   G   R   D   S   A   A   -
     CGGAGGTCCCCGATGACCGGTGGGTCCCCGACGAGCTGATGGCCTCCGACGCGGGGGG

6781 --------+---------+---------+---------+---------+---------+ 6840
          E   V   P   D   D   R   G   W   V   P   D   E   L   M   A   S   D   A   A   G   -
     GAACCCGCGCCCACGGCAACTTCATGGGGGCCGGTGACTTCGACGCGGCGGTTCTTCG

6841 --------+---------+---------+---------+---------+---------+ 6900
          T   R   A   H   G   N   F   M   A   G   A   G   D   F   D   A   A   F   F   G   -
```

```
6901 GGATCTCGCCGCGCGGAGGCGCTGGCGATGGACCCGCAGCGCCAGGCGCTGGAGACGA 6960
     ------+---------+---------+---------+---------+---------+
         I  S  P  R  E  A  L  A  M  D  P  Q  Q  R  Q  A  L  E  T  T  -

6961 CGTGGGAGGCGCTGGAAAGCGGCGGCATCCCACCGGAGACGTTGCGCGGCAGCGACACCG 7020
     ------+---------+---------+---------+---------+---------+
         W  E  A  L  E  S  A  G  I  P  P  E  T  L  R  G  S  D  T  G  -

7021 GCGTGTTCGTCGGCATGTCCCACCAGGGCTACGCGACCGGCCGCCCCGGAGGACG 7080
     ------+---------+---------+---------+---------+---------+
         V  F  V  G  M  S  H  Q  G  Y  A  T  G  R  P  R  P  E  D  G  -

7081 GCGTCGACGGGTACCTGCTCACCGGCAACACCGCGAGCGTCGGGACGCATCGCCT 7140
     ------+---------+---------+---------+---------+---------+
         V  D  G  Y  L  L  T  G  N  T  A  S  V  A  S  G  R  I  A  Y  -

7141 ACGTGCTGGGGCTGGAAGGTCCCGCGCTGACGGTGGACACGGCGTGTTCGTCGTTGG 7200
     ------+---------+---------+---------+---------+---------+
         V  L  G  L  E  G  P  A  L  T  V  D  T  A  C  S  S  S  L  V  -

7201 TGGCGTTGCACACGGGCGTGTGGGGTCGTTGCGTGACTGGGTGTCTTGGCGGTGGCCG 7260
     ------+---------+---------+---------+---------+---------+
         A  L  H  T  A  C  G  S  L  R  D  G  D  C  G  L  A  V  A  G  -
```

FIG.2-18

FIG.2-19

```
7261 GTGGTGTGTCGGTGATGGCGGGTCCGGAGGTGTTCACCGAGTTCTCCGCCAGGGCGCGC 7320
     ----+----+----+----+----+----+----+----+----+----+----+----+
      G  V  S  V  M  A  G  P  E  V  F  T  E  F  S  R  Q  G  A  L  -

7321 TCTCGCCGGACGGCCGGTGCAAGCCCTTCTCGGACGAGGCCGACGGATTCGGTCTCGGGG 7380
     ----+----+----+----+----+----+----+----+----+----+----+----+
      S  P  D  G  R  C  K  P  F  S  D  E  A  D  G  F  G  L  G  E  -

7381 AGGGTTCGGGCGTTCGTGCTCGTGCTCCAGCGGTTGTCCGACGCCAGGCGGGAGGGCCGCCG 7440
     ----+----+----+----+----+----+----+----+----+----+----+----+
      G  S  A  F  V  V  L  Q  R  L  S  D  A  R  R  E  G  R  R  V  -

7441 TGCTCGGCGTGGTGGCCGGGTCCGCGGTGAACCAGGACGGCGCGAGCAACGGGCTCTCCG 7500
     ----+----+----+----+----+----+----+----+----+----+----+----+
      L  G  V  V  A  G  S  A  V  N  Q  D  G  A  S  N  G  L  S  A  -

7501 CTCCGAGCGGCGTCGCGCAGCAGCGGGTCATCCGCCGGGCGTGGGCGCGTGCGGGGATCA 7560
     ----+----+----+----+----+----+----+----+----+----+----+----+
      P  G  V  A  Q  Q  R  V  I  R  R  A  W  A  R  A  G  I  T  -

7561 CGGGCGCGGGATGTGTGGCCGTGGTGGAGGCGCATGGGACCGGTACGCGGCTGGGCGATCCGG 7620
     ----+----+----+----+----+----+----+----+----+----+----+----+
      G  A  D  V  A  V  V  E  A  H  G  T  G  T  R  L  G  D  P  V  -
```

```
7621 TGGAGGCGTCGGCGTTGCTGGCTACTTACGGCAAGTCGCGCGGGTCGTCGGGCCCGGTGC
     --------+---------+---------+---------+---------+---------+ 7680
        E  A  S  A  L  L  A  T  Y  G  K  S  R  G  S  S  G  P  V  L -

7681 TGCTGGGTTCGGTGAAGTCGAACATCGGTCACGCGCAGGCGGCCCGGGGTGTCGCGGGGCG
     --------+---------+---------+---------+---------+---------+ 7740
        L  G  S  V  K  S  N  I  G  H  A  Q  A  A  A  G  V  A  G  V -

7741 TGATCAAGGTGCTCGGCCTGGAACGCGGGGTGGTGCCCCCGATGCTGTGCCGGGGGCG
     --------+---------+---------+---------+---------+---------+ 7800
        I  K  V  L  L  G  L  E  R  G  V  V  P  P  M  L  C  R  G  E -

7801 AGAGGTCGGGCCTCATCGACTGGTCCTCCGGGCGAGATCGAGCTCGCAGACGGCGTGCGGG
     --------+---------+---------+---------+---------+---------+ 7860
        R  S  G  L  I  D  W  S  S  G  E  I  E  L  A  D  G  V  R  E -

7861 AGTGGTCGCCCGCCGACGGGGTGCGCCGGGCAGGTGTGTCGGGCGTTCGGGGTGAGCG
     --------+---------+---------+---------+---------+---------+ 7920
        W  S  P  A  A  D  G  V  R  R  A  G  V  S  A  F  G  V  S  G -

7921 GGACGAACGCGCACGTGATCATCGCCCGGAGCCCGTGCCGCAACCGC
     --------+---------+---------+---------+---------+---------+ 7980
```

FIG.2-20

```
       T  N  A  H  V  I  I  A  E  P  P  E  P  E  P  V  P  Q  P  R  -
7981   GACGGCATGCTGCCCGCGACCGGGGGTGGTGCCGGTCGTGTCGGCCAGGACCGGGGCGG
       ----+---------+---------+---------+---------+---------+  8040

       R  M  L  P  A  T  G  V  V  P  V  V  L  S  A  R  T  G  A  A  -
8041   CGTTGCGGGCGGCAGGCCGGCTCGCCGACCACCTCGCCGCGCATCCCGGGATCGCGCAC
       ----+---------+---------+---------+---------+---------+  8100

       L  R  Q  A  G  R  L  A  D  H  L  A  A  H  P  G  I  A  P  -
8101   CGGCCGACGTGAGCTGGACGATGGCGGCCGCCAGCACTTCGAGGAGCGGGGCCGGCGG
       ----+---------+---------+---------+---------+---------+  8160

       A  D  V  S  W  T  M  A  R  A  R  Q  H  F  E  E  R  A  A  V  -
8161   TGCTCGCGCCGCCGACACCGGCCGGTGCCGTGCACCGGTTGGCGGGTGGCCGACGGCGGCGG
       ----+---------+---------+---------+---------+---------+  8220

       L  A  A  D  T  A  E  A  V  H  R  R  L  R  A  V  A  D  G  A  V  -
8221   TGGTTCCCGGTGTTGTCACCGGCCAGTGCCTCCGACGGTGGTTCAGTGTTCGTCTTCCCTG
       ----+---------+---------+---------+---------+---------+  8280

       V  P  G  V  V  T  G  S  A  S  D  G  G  S  V  F  V  F  P  G  -
8281   GGCAGGGTGCCCAGTGGGAAGGCATGGCGCGCGGGGAGTTGTTGCCGGTTCCCGTCTTCGCCG
       ----+---------+---------+---------+---------+---------+  8340
```

FIG.2-21

FIG.2-22

```
8641  GGGGGAGCATGCTCTCGGTGCGGGGCCGCTCCGACGTCGAGAAGCTGCTCGCCGACG  8700
      ------+---------+---------+---------+---------+---------+
       G   S   M   L   S   V   R   G   G   R   S   D   V   E   K   L   L   A   D   D  -

8701  ACAGCTGGACCGGGCAGGCTGGAGGTCGCGCGGTCAACGCCCCGACGCCGTGGTGG  8760
      ------+---------+---------+---------+---------+---------+
       S   W   T   G   R   L   E   V   A   A   V   N   G   P   D   A   V   V   V   A  -

8761  CCGGTGACGCCCCAGGCGGCGAGTTCCTGGAGTACTGCGAGGGCGTGGGCATCCGCG  8820
      ------+---------+---------+---------+---------+---------+
       G   D   A   Q   A   A   R   E   F   L   E   Y   C   E   G   V   G   I   R   A  -

8821  CCCGGCGCGATCCCGGTGGACTACGCCCTCGCACACCGGCACGTCGAGCCCGTGCGGCGACG  8880
      ------+---------+---------+---------+---------+---------+
       R   A   I   P   V   D   Y   A   S   H   T   A   H   V   E   P   V   R   D   E  -

8881  AACTGGTCCAGGCGCTGGCCGGGATCACCCCGCGACGGGCCGAGGTGCCGTTCTTCTCCA  8940
      ------+---------+---------+---------+---------+---------+
       L   V   Q   A   L   A   G   I   T   P   R   R   A   E   V   P   F   F   S   T  -

8941  CCCTGACCGGGCGACTTCCTCGACGGCACCGAGCTGGACGCGGGCTACTGGTACCGCAACC  9000
      ------+---------+---------+---------+---------+---------+
       L   T   G   D   F   L   D   G   T   E   L   D   A   G   Y   W   Y   R   N   L  -
```

FIG.2-23

```
9001  TGCGTCACCCGGTGGAGTTCCACTCCGCGCCGTGCCGGCGCTGACCGACCAGGGATACGCGA
      ---------+---------+---------+---------+---------+---------+  9060
       R  H  P  V  E  F  H  S  A  V  Q  A  L  T  D  Q  G  Y  A  T  -

9061  CGTTCATCGAGGTCAGCCCGCACCCGGTGCTGGCGTCCAGCGTCCAGGAGACCCTCGACG
      ---------+---------+---------+---------+---------+---------+  9120
       F  I  E  V  S  P  H  P  V  L  A  S  S  V  Q  E  T  L  D  D  -

9121  ACGCCGAGTCGGACGCGGCGGTGCTCGGGACGCTGGAACGCGACGCGGGCGACGCCGACC
      ---------+---------+---------+---------+---------+---------+  9180
       A  E  S  D  A  A  V  L  G  T  L  E  R  D  A  G  D  A  D  R  -

9181  GCTTCCTCACGGCACTCGCCGACGCGCACACGCGCGGTGTCGCGGTCGACTGGGAAGCGG
      ---------+---------+---------+---------+---------+---------+  9240
       F  L  T  A  L  A  D  A  H  T  R  G  V  A  V  D  W  E  A  V  -

9241  TGCTCGGCGGCCGGACTGGTCGACCTGCCGGGGTTATCCTTTCCAGGGCAAGCGGGTTCT
      ---------+---------+---------+---------+---------+---------+  9300
       L  G  R  A  G  L  V  D  L  P  G  Y  P  F  Q  Q  G  K  R  F  W  -

9301  GGCTGCTGCCGGACCGCACCCCCTCGTGACGAGCTCGACGGCTGGTTCTACCGGGGTCG
      ---------+---------+---------+---------+---------+---------+  9360
       L
```

FIG.2-24

```
          L   L   P   D   R   T   T   P   R   D   E   L   D   G   W   F   Y   R   V   D   -
9361  ACTGGACCGAGGTGCCGGCTCCGAACCTGCCGGCGCTGCCGGCCGTTGGCTCGTGGTGG
      ---------+---------+---------+---------+---------+---------+  9420

          W   T   E   V   P   R   S   E   P   A   A   L   R   G   R   W   L   V   V   V   -
9421  TGCCCGAGGGGCACGAGGAGGACGGCTGGACCGTCGAGGTGCCGCGTCCGCCGAGG
      ---------+---------+---------+---------+---------+---------+  9480

          P   E   G   H   E   E   D   G   W   T   V   E   V   R   S   A   L   A   E   A   -
9481  CCGGCGCCGAACCGGAGGTCACGCGCGGCCGTCGGCGGCTGTGTCGGTGACTGCGCGGGGCG
      ---------+---------+---------+---------+---------+---------+  9540

          G   A   E   P   E   V   T   R   G   V   G   G   L   V   G   D   C   A   G   V   -
9541  TGGTGTCGTTGCTCGCCCTCGAGGGGCGATGGTGCGCAAACCCTTGTGCTGGTGCGGG
      ---------+---------+---------+---------+---------+---------+  9600

          V   S   L   L   A   L   E   G   D   G   A   V   Q   T   L   V   L   V   R   E   -
9601  AACTCGACGCCGAGGGCCATCGACGCGCCACTGTGGACGGTCACCTTCGGGCGGTCGACG
      ---------+---------+---------+---------+---------+---------+  9660

          L   D   A   E   G   I   D   A   P   L   W   T   V   T   F   G   A   V   D   A   -
9661  CGGGCAGTCCGGTGGCCCGGACCAGGCGAAGCTGTGGGGGCTGGGGCCAGTCGCGT
      ---------+---------+---------+---------+---------+---------+  9720
```

FIG.2-25

FIG.2-26

```
        G  S  P  V  A  R  P  D  Q  A  K  L  W  G  L  G  Q  V  A  S  -
9721   CCCTGGAACGCGGGCCCCGCTGGACCGGGCCTCGTCGGCACATGCCGACCCGG
       ---------+---------+---------+---------+---------+---------+ 9780

        L  E  R  G  P  R  W  T  G  L  V  D  L  P  H  M  P  D  P  E  -
9781   AACTGCGAGGCCGTCTCACCGCGGTGCCGGTGCCGGAGGACCAGGTCGCGGTGCGCG
       ---------+---------+---------+---------+---------+---------+ 9840

        L  R  G  R  L  T  A  V  L  A  G  S  E  D  Q  V  A  V  R  A  -
9841   CCGACGCCGGTGCCGTGGCGGCGGCTTCCCCGCCCACGTCACCGCCACCTCGGAGTACG
       ---------+---------+---------+---------+---------+---------+ 9900

        D  A  V  R  A  R  R  L  S  P  A  H  V  T  A  T  S  E  Y  A  -
9901   CGGTGCCGGGCGGGCACAATCCTGGTCACCGGTGGCACCGGCCCGGCGCCTGGGCGGAGGTGG
       ---------+---------+---------+---------+---------+---------+ 9960

        V  P  G  G  T  I  L  V  T  G  G  T  A  G  L  G  A  E  V  A  -
9961   CCCGGTGGCTCGCCGGTCGCCGGTCGCCGAACACCTCGCTGGTCAGCAGGCCGAGGCCCGG
       ---------+---------+---------+---------+---------+---------+ 10020

        R  W  L  A  G  R  G  A  E  H  L  A  L  V  S  R  R  G  P  D  -
       ACACCGAGGGCGTCGGCCGACCTGACCGCGCCGAGCTGACCCGGCTCGGCGGGGGTGTCGG
```

FIG.2-27

```
10021 ------+---------+---------+---------+---------+---------+ 10080
          TGCACGCGGTGCGACGTCAGCAGCCGCGAACCGGTGAGGAACTCGTGCACGGCCTGATCG
            T  E  G  V  G  D  L  T  A  E  L  T  R  L  G  A  R  V  S  V

10081 ------+---------+---------+---------+---------+---------+ 10140
          AGCAGGGGCGACGTCGTCCGCGGTGTGGTGCACGCGGGGACTGCCGCGCAGCAGGTCGCGA
            H  A  C  D  V  S  S  R  E  P  V  R  E  L  V  H  G  L  I  E

10141 ------+---------+---------+---------+---------+---------+ 10200
          TCAACGACATGGACGAGGCCGCCTTCGACGAGGTGGTCGCGGCCAAGGCCGGGGGCGCGG
            Q  G  D  V  V  R  G  V  V  H  A  A  G  L  P  Q  Q  V  A  I

10201 ------+---------+---------+---------+---------+---------+ 10260
          TGCACCTGGACGAGCTGTGCTCGGACGCGCCGAGCTGTTCCTGCTGTTCCTCCCGGGGCCG
            N  D  M  D  E  A  A  F  D  E  V  V  A  A  K  A  G  G  A  V

10261 ------+---------+---------+---------+---------+---------+ 10320
          GGGTGTGGGGGAAGCGCCCCGGCCTACGCGCCGGCAACGCGTTCCTGGACGCGCCT
            H  L  D  E  L  C  S  D  A  E  L  F  L  F  S  S  G  A  G

10321 ------+---------+---------+---------+---------+---------+ 10380
            V  W  G  S  A  R  Q  G  A  Y  A  A  G  N  A  F  L  D  A  F
```

FIG.2-28

```
10381  TCGCCCGGCACCGCCGGGGGCCGGCCTGCCCGCCACGTCGTTGGGGGCTGTGGG
       ---------+---------+---------+---------+---------+---------+  10440
             A  R  H  R  R  G  R  G  L  P  A  T  S  V  A  W  G  L  W  A   -

10441  CGGCGGGGGCATGACCGGCGACGAGGAGGCCGTGTCGTTCCTGCGCGAGCGCGGTGTGC
       ---------+---------+---------+---------+---------+---------+  10500
             A  G  G  M  T  G  D  E  E  A  V  S  F  L  R  E  R  G  V  R   -

10501  GGGCGATGCCCGTACCGCGCGCCCTCGCGCCCTGGACAGGGTGCTGGCCTCCGGGGAGA
       ---------+---------+---------+---------+---------+---------+  10560
             A  M  P  V  P  R  A  L  A  A  L  D  R  V  L  A  S  G  E  T   -

10561  CGGCGGTGGTCGTGACGGACGTGGACTGGCCCGCCTTCGCGGAGTCCTACACGCCGCCC
       ---------+---------+---------+---------+---------+---------+  10620
             A  V  V  V  T  D  V  D  W  P  A  F  A  E  S  Y  T  A  A  R   -

10621  GGCCCCGGCCCGTTGCTCGACCGCATCGTCACGACGGCGCCGAGCGAGCGGGCCGGAGAAC
       ---------+---------+---------+---------+---------+---------+  10680
             P  R  P  L  L  D  R  I  V  T  T  A  P  S  E  R  A  G  E  P   -

10681  CGGAGACGGAGAGCCTGCGCGACCGGCGTCTGCCCGAGCCGTGCCGAGCGGACGGCGG
       ---------+---------+---------+---------+---------+---------+  10740
             E  T  E  S  L  R  D  R  L  A  G  L  P  R  A  E  R  T  A  E   -
```

```
10741  AGCTGGTGGCGCCTGGTCCGCCAGCACCGGCGACCGTGCTGGGCCACGACGACCCGAAGG
       ------+---------+---------+---------+---------+---------+  10800
         L  V  R  L  V  R  T  S  T  A  T  V  L  G  H  D  D  P  K  A  -

10801  CGGTGGCGGCGACCACGCCGTTCAAGGAGCTCGGGTTCGGCGGCCGTCCGGC
       ------+---------+---------+---------+---------+---------+  10860
         V  R  A  T  T  P  F  K  E  L  G  F  D  S  L  A  A  V  R  L  -

10861  TGCGCAACCTGCTCAACGCGGGCCTCCGCCTGCCGTCGACGCTGGTCTTCGACC
       ------+---------+---------+---------+---------+---------+  10920
         R  N  L  L  N  A  A  T  G  L  R  L  P  S  T  L  V  F  D  H  -

10921  ACCCGAACGCCTCCGCGGTCGCCGGGTTCCTCGACGCCGAGCTCGGCACCGAGGTCCGGG
       ------+---------+---------+---------+---------+---------+  10980
         P  N  A  S  A  V  A  G  F  L  D  A  E  L  G  T  E  V  R  G  -

10981  GGGAGGCGCCGTCGCGGGCCCTCGCGGGCTGGACGCGCTGGAAGGGCCCTGCCCGAGGTGC
       ------+---------+---------+---------+---------+---------+  11040
         E  A  P  S  A  L  A  G  L  D  A  L  E  G  A  L  P  E  V  P  -

11041  CCGCAACCGAGCGGGAAGAGCTGGTACAGCGGCTTGGAACGGATGCTCGCCGCGCTACGCC
       ------+---------+---------+---------+---------+---------+  11100
         A  T  E  R  E  E  L  V  Q  R  L  E  R  M  L  A  A  L  R  P  -
```

FIG.2-29

```
11101  CGGTCGCGCCAGGCCGGCCGACGCCTCCGGGCCGCCAACCCGTCCGGCGACGACCTGG  11160
       ---------+---------+---------+---------+---------+---------+
         V  Q  A  A  D  A  S  G  T  G  A  N  P  S  G  D  D  L  G  -

11161  GCGAGGCGGGCGTGGACGAACTGCTCGAAGCACTCGGCCGGGAGCTCGACGGCGATTGA  11219
       ---------+---------+---------+---------+---------+---------+
         E  A  G  V  D  E  L  L  E  A  L  G  R  E  L  D  G  D  *

12643  CCGCCGATTGGAGAAAAGGTGACTGACAGCGGAGAAGGTGGCGGAGTACCTCCGTCGGGCG  12702
       ---------+---------+---------+---------+---------+---------+
              V  T  D  S  E  K  V  A  E  Y  L  R  R  A  -

12703  ACGCTCGACCTGCCGTGCCGCCCGGCGCATCCCGGCAGCGCTGGAATCCGACCCGATCGCC  12762
       ---------+---------+---------+---------+---------+---------+
         T  L  D  L  R  A  A  R  Q  R  I  R  E  L  E  S  D  P  I  A  -

12763  ATCGTCAGCATGGCCTGCCGCCTGCCGGGGGGTGAACACACCCCCAGCGGCTGTGGGAG  12822
       ---------+---------+---------+---------+---------+---------+
         I  V  S  M  A  C  R  L  P  G  G  V  N  T  P  Q  R  L  W  E  -
```

FIG.2-30

FIG.2-31

```
12823 ------+---------+---------+---------+---------+---------+ 12882
           L  L  R  E  G  G  E  T  L  S  G  F  P  T  D  R  G  W  D  L
      GCGGGGCTGCACCACCCCGACCCCGGTACCAGCTACGTCGACAAGGGCGGG

12883 ------+---------+---------+---------+---------+---------+ 12942
           A  R  L  H  H  P  D  P  D  N  P  G  T  S  Y  V  D  K  G  G
      TTCCTCGACGACGCGGGGGGCTTCGACGCGGGAGTTCTTCGGCGTCCGCGGCCGAGGCC

12943 ------+---------+---------+---------+---------+---------+ 13002
           F  L  D  D  A  A  G  F  D  A  E  F  F  G  V  S  P  R  E  A
      GCGGCCATGGACCCGCAGCAGCGGCTGCTGGAGACGAGCTGGTGGAGAAC

13003 ------+---------+---------+---------+---------+---------+ 13062
           A  A  M  D  P  Q  Q  R  L  L  E  T  S  W  E  L  V  E  N
      GCCGGCATCGACCCGCACTCGCTGCGCGGTACCGCGACCGGCGTCTTCCTCGGAGTGGCG

13063 ------+---------+---------+---------+---------+---------+ 13122
           A  G  I  D  P  H  S  L  R  G  T  A  T  G  V  F  L  G  V  A
      AAGTTCGGCTACGGCGAGGACACCGCCGCGGAGGACGTCGAGGGCTACTCGGTCACC

13123 ------+---------+---------+---------+---------+---------+ 13182
           K  F  G  Y  G  E  D  T  A  A  A  E  D  V  E  G  Y  S  V  T
```

FIG.2-32

```
13183 GGTGTGGCGCCCGGCCGGTCGCCTCCGGCGCATCTCCTACACCATGGGCCTGGAGGGGCCG 13242
      -----+---------+---------+---------+---------+---------+
       G  V  A  P  A  V  A  S  G  R  I  S  Y  T  M  G  L  E  G  P

13243 TCGATCAGCGTCGACACCGCGTGCTCGTCGCTGGTCGCGCTGCACCTGGCGGTCGAG 13302
      -----+---------+---------+---------+---------+---------+
       S  I  S  V  D  T  A  C  S  S  S  L  V  A  L  H  L  A  V  E

13303 TCGCTGCGCAAGGGCGAGTCGTCGATGGCGGTCGTGGGCGGTGCGGTGATGGCGACC 13362
      -----+---------+---------+---------+---------+---------+
       S  L  R  K  G  E  S  S  M  A  V  V  G  G  A  A  V  M  A  T

13363 CCGGGGGTGTTCGTCGACTTCAGCCGGCAGCGCGCTCGCGCCGCCGACGGCGGTCGAAG 13422
      -----+---------+---------+---------+---------+---------+
       P  G  V  F  V  D  F  S  R  Q  R  A  L  A  A  D  G  R  S  K

13423 GCGTTCGGTGCCGGCGCCGACGGGTTCGGCTTCTCCGAAGGCGTCACCCTGGTCCTGCTC 13482
      -----+---------+---------+---------+---------+---------+
       A  F  G  A  G  A  D  G  F  G  F  S  E  G  V  T  L  V  L  L

13483 GAGCGGCTGTCGGAGGCGCGGCGCAACGGGCACGAGGTGCTGGCGGTTCGGCGGCTCG 13542
      -----+---------+---------+---------+---------+---------+
       E  R  L  S  E  A  R  R  N  G  H  E  V  L  A  V  V  R  G  S
```

FIG.2-33

```
13543 GCGCTCAACCAGGACGGGCCAGCAACGGGGCTTTCCGCGCCGAGCGGGCCCGCGCCAGCGC 13602
       ---------+---------+---------+---------+---------+---------+
        A  L  N  Q  D  G  A  S  N  G  L  S  A  P  S  G  P  A  Q  R

13603 AGGGTCATCCGGCAGGCCCTCGAGAGCTGCGGTCTGGAGCCCGGCGACGTCGACGCGGTG 13662
       ---------+---------+---------+---------+---------+---------+
        R  V  I  R  Q  A  L  E  S  C  G  L  E  P  G  D  V  D  A  V

13663 GAGGCGCACGGCACCGGTACGGCGCTCGGCGACCCGATCGAGGCGAACGCGCTGCTGGAC 13722
       ---------+---------+---------+---------+---------+---------+
        E  A  H  G  T  G  T  A  L  G  D  P  I  E  A  N  A  L  L  D

13723 ACCTACGGCCGCGACCGCGACGCCGACCGGCCGCTCTGGCTGGGTGAAGTCCAAC 13782
       ---------+---------+---------+---------+---------+---------+
        T  Y  G  R  D  R  D  A  D  R  P  L  W  L  G  S  V  K  S  N

13783 ATCGGCCACACCCAGGCGGCAGCGGGCGTCACCGGCGTCCTGAAGGTGGTCCTGGCGCTG 13842
       ---------+---------+---------+---------+---------+---------+
        I  G  H  T  Q  A  A  A  G  V  T  G  L  L  K  V  V  L  A  L

13843 CGCAACGGGGAACTGCCCGCGACCCTGCACGTCGAGGAGCCCACGCCGCACGTCGACTGG 13902
       ---------+---------+---------+---------+---------+---------+
        R  N  G  E  L  P  A  T  L  H  V  E  E  P  T  P  H  V  D  W
```

```
13903  TCGTCCGGCGGCCGTGGCCGTGGCCGCTGCTGCGGGCAACCAGCCGTGGCGGCGGGGCGAGCGGGACT  13962
       ------+---------+---------+---------+---------+---------+
        S  S  G  G  V  A  L  L  A  G  N  Q  P  W  R  R  G  E  R  T

13963  CGGCGGCGCCCGTGTTCCGCGTTCGGGGATCAGCGGGGACGAATGCGCACGTGATCGTCGAG  14022
       ------+---------+---------+---------+---------+---------+
        R  R  A  R  V  S  A  F  G  I  S  G  T  N  A  H  V  I  V  E

14023  GAAGCTCCTGAGCGGCGAGCACCGGGAGACCACCGCGCACGGACCGGTTCCGCTG  14082
       ------+---------+---------+---------+---------+---------+
        E  A  P  E  R  E  H  R  E  T  T  A  H  D  G  R  P  V  P  L

14083  GTGGTGTCCGCGCGCACGACGGCGGCGTTGCGCGGGCCCAGATCGCCGAGCTG  14142
       ------+---------+---------+---------+---------+---------+
        V  V  S  A  R  T  T  A  A  L  R  A  Q  A  A  A  Q  I  A  E  L

14143  CTCGAACGCCCGGACGCCGACCTCGCCGGGGTCGGGCCTGGCCTGGCCACGACCCGCGCC  14202
       ------+---------+---------+---------+---------+---------+
        L  E  R  P  D  A  D  L  A  G  V  G  L  G  L  A  T  T  R  A

14203  CGCCACGAGCACCGCGCCGTGGTGGCATCGACCCGGAAGCGGTGCGGGACTG  14262
       ------+---------+---------+---------+---------+---------+
```

FIG.2-34

```
R   H   E   H   R   A   A   V   V   A   S   T   R   E   E   A   V   R   G   L   -
      CGGGAGATCGCCGCCGGTGCCGCGACGGCCGACGCCGTGGTCGAGGGCGTCACCGAGGTG
14263 ---------+---------+---------+---------+---------+---------+   14322

R   E   I   A   A   G   A   A   T   A   D   A   V   V   E   G   V   T   E   V   -
      GACGGGGCGCAACGTCGTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGT
14323 ---------+---------+---------+---------+---------+---------+   14382

D   G   R   N   V   V   F   L   F   P   G   Q   G   S   Q   W   A   G   M   G   -
      GCCGAGCTGCTGTCGTCGTCGCCGGTGTTCGCCGGGGAAGATCCGGGCCTGCGACGAGTCG
14383 ---------+---------+---------+---------+---------+---------+   14442

A   E   L   L   S   S   S   P   V   F   A   G   K   I   R   A   C   D   E   S   -
      ATGGCCCCGATGCAGGAGGACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCCG
14443 ---------+---------+---------+---------+---------+---------+   14502

M   A   P   M   Q   D   W   K   V   S   D   V   L   R   Q   A   P   G   A   P   -
      GGCCTGGACCGGGTCGACGTGGTGCAGCCGGTGTTGTTCGCGGGTGATGGTGTCGCTGGCG
14503 ---------+---------+---------+---------+---------+---------+   14562

G   L   D   R   V   D   V   V   Q   P   V   L   F   A   V   M   V   S   L   A   -
      GAGCTGTGGCGCTCGTACGGCGTGGAGCCCGGCGTCGTGGGCACTCGCAGGGCGAG
14563 ---------+---------+---------+---------+---------+---------+   14622
```

FIG.2-35

FIG.2-36

```
          E   L   W   R   S   Y   G   V   E   P   A   A   V   V   G   H   S   Q   G   E   -
          ATCGGCCGGCCGCGCACGTCGCCGGGGGCGCTCACGTTGGAGGACGCGGGCGAAGCTCGTG
14623     ---------+---------+---------+---------+---------+---------+     14682

          I   A   A   A   H   V   A   G   A   L   T   L   E   D   A   A   K   L   V   V   -
          GGCCGCAGCCGCCTGATGCGGTCGCGCTCTCCGGGGAGGGGCATGGCCGCCGTCGCGCTG
14683     ---------+---------+---------+---------+---------+---------+     14742

          G   R   S   R   L   M   R   S   L   S   G   E   G   G   M   A   A   V   A   L   -
          GGCGAGGCCGCCGTGCGCGAGCGCCTGCGCCCGTGGCAGGACCGGCTCTCGGTGGCCGCG
14743     ---------+---------+---------+---------+---------+---------+     14802

          G   E   A   A   V   R   E   R   L   R   P   W   Q   D   R   L   S   V   A   A   -
          GTCAACGGTCCCCGGTCGGTGTCTCCGGCGCCCGGAGCCCGGCGCTGGCGTTTTCC
14803     ---------+---------+---------+---------+---------+---------+     14862

          V   N   G   P   R   S   V   V   V   S   G   E   P   G   A   L   R   A   F   S   -
          GAGGACTGCGCGGGGCCGAGGGCATCCGCGTCCGGCGACATCGACGTGGACTACGCCTCGCAC
14863     ---------+---------+---------+---------+---------+---------+     14922

          E   D   C   A   A   E   G   I   R   V   R   D   I   D   V   D   Y   A   S   H   -
```

```
14923 TCGCCGCAGATCGAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCG 14982
      ---------+---------+---------+---------+---------+---------+
        S   P   Q   I   E   R   V   R   E   E   L   L   E   T   T   G   D   I   A   P

14983 CGCCCGGCGCGGGTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAG 15042
      ---------+---------+---------+---------+---------+---------+
        R   P   A   R   V   T   F   H   S   T   V   E   S   R   S   M   D   G   T   E

15043 CTGGATGCCCGGTACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTG 15102
      ---------+---------+---------+---------+---------+---------+
        L   D   A   R   Y   W   Y   R   N   L   R   E   T   V   R   F   A   D   A   V

15103 ACGCGGCTGGCCGAGTCGGGATACGACGCGGTTCATCGAGGTCAGCGGCCATCCGGTCGTG 15162
      ---------+---------+---------+---------+---------+---------+
        T   R   L   A   E   S   G   Y   D   A   F   I   E   V   S   P   H   P   V   V

15163 GTCCAGGCCGTCGAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCGTAGTC 15222
      ---------+---------+---------+---------+---------+---------+
        V   Q   A   V   E   E   A   V   E   E   A   D   G   A   E   D   A   V   V   V

15223 GGCTCGCTGCACCGGCGACGGCGGGTGACCTCTCCTGCGGCTTCCTGCGGTCGATGGCCACCGCG 15282
      ---------+---------+---------+---------+---------+---------+
        G   S   L   H   R   D   G   G   D   L   S   A   F   L   R   S   M   A   T   A
```

FIG.2-37

```
15283 CACGGTGTCCGGTGTGTGGACATCAGGTGGGACGTCGCTCTGCCCGGCCGGGCGCCCTTCGCG 15342
      ------+---------+---------+---------+---------+---------+--
       H  V  S  G  V  D  I  R  W  D  V  A  L  P  G  A  A  P  F  A  -

15343 CTGCCGACGTATCCGTTCCAGCGCCAAGCGCTACTGGCTCCAGCCCGCCACCCGCGCC 15402
      ------+---------+---------+---------+---------+---------+--
       L  P  T  Y  P  F  Q  R  K  R  Y  W  L  Q  P  A  A  P  A  A  -

15403 GCCTCCGACGAGCTGGCCTACCGCGTTCCTGGACTCCGATCGAAAAGCCGGAGTCGGGA 15462
      ------+---------+---------+---------+---------+---------+--
       A  S  D  E  L  A  Y  R  V  S  W  T  P  I  E  K  P  E  S  G  -

15463 AACCTGGACGGCGACTGGTTGGTTGTCACACCCCTCATCAGTCCGGAGTGGACGGAAATG 15522
      ------+---------+---------+---------+---------+---------+--
       N  L  D  G  D  W  L  V  V  T  P  L  I  S  P  E  W  T  E  M  -

15523 CTGTGCGAGGCCATCAACGCCAACGGTGCCAGGGCCGTTGCGCTGCGAGGTGGACACGTCC 15582
      ------+---------+---------+---------+---------+---------+--
       L  C  E  A  I  N  A  N  G  G  R  A  L  R  C  E  V  D  T  S  -

15583 GCTTCGCGCACTGAGATGCCCCAGGCCGTCGCACAGGCCGAACGGGATTCCGGGGGCGTG 15642
      ------+---------+---------+---------+---------+---------+--
       L                                                         -
```

FIG.2—38

```
      A  S  R  T  E  M  A  Q  A  V  A  Q  A  G  T  G  F  R  G  V   -
          CTCTCGTTGCTGTCGTCGGACGAATCCGGCCTGCCGTCCGGGGGTTCCTGCCGTGCGGTC
15643     ---------+---------+---------+---------+---------+---------+   15702

      L  S  L  L  S  S  D  E  S  A  C  R  P  G  V  P  A  G  A  V   -
          GGCCTGCTCACCCTGGTCCAGGCGCTGGGCGATGCCGGGGTCGACGCACCGGTGTGGTGC
15703     ---------+---------+---------+---------+---------+---------+   15762

      G  L  L  T  L  V  Q  A  L  G  D  A  G  V  D  A  P  V  W  C   -
          CTGACCCCAGGGTGCGGTCCGCCACTCCCGCCGACGACGACCTCGCCGCCTGCCGCAGACC
15763     ---------+---------+---------+---------+---------+---------+   15822

      L  T  Q  G  A  V  R  T  P  A  D  D  D  D  L  A  R  P  A  Q  T   -
          ACCGCGGCACGGCTTCGCGCCAGGTCGCCAGGTCGCTGGAGCTGCCGGGGCGTGGGGCGGTGTG
15823     ---------+---------+---------+---------+---------+---------+   15882

      T  A  H  G  F  A  Q  V  A  G  L  E  L  P  G  R  W  G  G  V   -
          GTCGACCTGCCCGAATCGGTCGACGACGCGGCGCTGCGTCTGCGTCTGCTCGTGGCAGTCCTGCGC
15883     ---------+---------+---------+---------+---------+---------+   15942

      V  D  L  P  E  S  V  D  D  A  A  L  R  L  L  V  A  V  L  R   -
          GGCGGCGGCCGTGCCGAGGACCACCTCGCCGGACGCCGCCTCCACGGCCGTCGC
15943     ---------+---------+---------+---------+---------+---------+   16002
```

FIG.2-39

```
        G  G  G  R  A  E  D  H  L  A  V  R  D  G  R  L  H  G  R  R
16003   GTCGTCCGGCGCAAGCCTGCCGCAGTCCGGCTCGGCGAGCTGGACCCCGCACGGGACCGTG   16062
        ---------+---------+---------+---------+---------+---------+

        V  V  R  A  S  L  P  Q  S  G  S  R  S  W  T  P  H  G  T  V
16063   CTGGTCACCGGCGGGGCGGGCCCCGTCGGGCGACCAACTGTGCGGTGCTCGCCGACCGG   16122
        ---------+---------+---------+---------+---------+---------+

        L  V  T  G  A  A  S  P  V  G  D  Q  L  V  R  W  L  A  D  R
16123   GGAGCCGGAGCGGGCTGGTGCTGGCCGGAGCCTGCCGACGACCTGCTGGCCGCGGTC   16182
        ---------+---------+---------+---------+---------+---------+

        G  A  E  R  L  V  L  A  G  A  C  P  G  D  D  L  L  A  A  V
16183   GAGGAAGCGGGGCGCATCGGCCGTCGTGTGCGCCCAGGACGCGGGCCTGCGCGGAGGCG   16242
        ---------+---------+---------+---------+---------+---------+

        E  E  A  G  A  S  A  V  V  C  A  Q  D  A  A  A  L  R  E  A
16243   CTCGGCGACGAGCCGGTGACCGCGCTCGTGCACCGCGGAACCCTGACGAACTTCGGCAGC   16302
        ---------+---------+---------+---------+---------+---------+

        L  G  D  E  P  V  T  A  L  V  H  A  G  T  L  T  N  F  G  S
        ATCAGCGAAGTCGCACCGGAGGAGTTCGCCGAGACGATCGGCGCCAAGACCGCGTTGCTC
```

FIG.2-40

FIG.2—41

```
16303 --------+---------+---------+---------+---------+---------+ 16362
           I  S  E  V  A  P  E  E  F  A  E  T  I  A  A  K  T  A  L  L
       GCCGTGCTGGACGAAGTCCTCGGGCGACCGGCCGTCGAGCGGGGAGGTCTACTGCTCGTCG
16363 --------+---------+---------+---------+---------+---------+ 16422
           A  V  L  D  E  V  L  G  D  R  A  V  E  R  E  V  Y  C  S  S
       GTCGCCGGGATCTGGGGCGGCGGGATGGCCGCCTACGGCGCAGGCGCCTACCTC
16423 --------+---------+---------+---------+---------+---------+ 16482
           V  A  G  I  W  G  G  A  G  M  A  A  Y  A  A  G  S  A  Y  L
       GACGCGCTGGCCGAGCACCACCGCGCGCGCTCGTGCACCTCGGTGCGCCTGGACG
16483 --------+---------+---------+---------+---------+---------+ 16542
           D  A  L  A  E  H  H  R  A  R  G  R  S  C  T  S  V  A  W  T
       CCGTGGGCGCTGCCGGGCGGTGCGGTGGACGACGGCTACCTGCGGGAACGCGGACTGCGC
16543 --------+---------+---------+---------+---------+---------+ 16602
           P  W  A  L  P  G  G  A  V  D  D  G  Y  L  R  E  R  G  L  R
       AGCCTCTCCGCCGACAGGGCGGATGCGCCACCTGGGAGCGGGTGCTGGCCGCCGGGGTG
16603 --------+---------+---------+---------+---------+---------+ 16662
           S  L  S  A  D  R  A  M  R  T  W  E  R  V  L  A  A  G  P  V
```

FIG.2-42

```
16663  TCGGTCGGTGGCCGACGTGGACTGGGG---+---------+---------+---------+---------+---------+  16722
       AGCCAGCCACCGGCTGCACCTGACCCC
        S  V  A  D  V  D  W  P  V  L  S  E  G  F  A  A  T  R  -

16723  CCGACCGGCTGTTCGCCGAACTCGCCGGGCGGGGACAGGCGGAGGCCGAGCCGGAC---+---------+---------+---------+---------+---------+  16782
        P  T  A  L  F  A  E  L  A  G  R  G  G  Q  A  E  A  E  P  D  -

16783  AGCGGGACCGGCGAGCCGGCACAACGGTCTCGCGGGCTTCCCCGACGAGCAGCAG---+---------+---------+---------+---------+---------+  16842
        S  G  P  T  G  E  P  A  Q  R  L  A  G  L  S  P  D  E  Q  Q  -

16843  GAAAACCTGCTCGAACTCGTCGCGAACGCGGTTGCCGAGGTGCTTGGCCACGAGTCCGCC---+---------+---------+---------+---------+---------+  16902
        E  N  L  L  E  L  V  A  N  A  V  A  E  V  L  G  H  E  S  A  -

16903  GCCGAGATCAACGTGCGCCGGTTCAGCGAGCTCGGACTCGATCGCTCAACGCGGATG---+---------+---------+---------+---------+---------+  16962
        A  E  I  N  V  R  R  A  F  S  E  L  G  L  D  S  L  N  A  M  -

16963  GCCCTGCGCAAGCGCCTGTCGGCGAGCACCGGCCTGCGCCGGTCGCTGGTGTTC---+---------+---------+---------+---------+---------+  17022
        A  L  R  K  R  L  S  A  S  T  G  L  R  L  P  A  S  L  V  F  -
```

17023 GACCACCCCACGTCACCGGCTCGGCGCCAGCACCTGCCGCCCGGCTCGTCGGTGACGCC 17082
　　　　　　----+----+----+----+----+----+----+----+----+----+
　　　　　D　H　P　T　V　T　A　L　A　Q　H　L　R　A　R　L　V　G　D　A

17083 GACCAGGCCGGCGGTGCGTCGGCCGCGACGAGTCCGAGCCCATCGCCATCGTC 17142
　　　　　　----+----+----+----+----+----+----+----+----+----+
　　　　　D　Q　A　A　V　R　V　V　G　A　A　D　E　S　E　P　I　A　I　V

17143 GGCATCGGCTGCCGTTCCCCGGCGCTCGCCCGAGCAGTTGTGGCGGGTGCTG 17202
　　　　　　----+----+----+----+----+----+----+----+----+----+
　　　　　G　I　G　C　R　F　P　G　G　I　G　S　P　E　Q　L　W　R　V　L

17203 GCCGAGGGCGCGCGAACCTCACCACCGGCTTCCCGGCGCCGACCGGCTGGGACATCGGGGCGG 17262
　　　　　　----+----+----+----+----+----+----+----+----+----+
　　　　　A　E　G　A　N　L　T　T　G　F　P　A　D　R　G　W　D　I　G　R

17263 CTCTACCACCCGGACCCGGACAACCCGGACAAGGGCGGGTTCCTC 17322
　　　　　　----+----+----+----+----+----+----+----+----+----+
　　　　　L　Y　H　P　D　P　D　N　P　G　T　S　Y　V　D　K　G　G　F　L

17323 ACCGACGCGGCGGATTCGACCCGGGCTTCTTCGGCATCACGCCCCGCGAAGGCGCTGGCG 17382
　　　　　　----+----+----+----+----+----+----+----+----+----+
　　　　　T　D　A　A　D　F　D　P　G　F　F　G　I　T　P　R　E　A　L　A

FIG.2-43

```
17383 ATGGACCCGCAGCAGCGCCTCATGCTGGAGACGGCGTGGGAGGCAGTGGAACGCGCGGGC 17442
      --------+---------+---------+---------+---------+---------+--
       M  D  P  Q  Q  R  L  M  L  E  T  A  W  E  A  V  E  R  A  G  -

17443 ATCGACCCCGACGCGCCTGCGAGGCACCGACACCGGCGTCTTCGTCGGCATGAACGGCCAG 17502
      --------+---------+---------+---------+---------+---------+--
       I  D  P  D  A  L  R  G  T  D  T  G  V  F  V  G  M  N  G  Q  -

17503 TCCTACATGCAGCTGCTGGCCGGTGAGGCCGAACGCGTCGACGGCTACCAGGGCCTCGGA 17562
      --------+---------+---------+---------+---------+---------+--
       S  Y  M  Q  L  L  A  G  E  A  E  R  V  D  G  Y  Q  G  L  G  -

17563 AACTCCGGCGAGCGTGCTCTCCGGGGCCATCGCCTACACCTTCGGCTGGGAGGGCCCGGCG 17622
      --------+---------+---------+---------+---------+---------+--
       N  S  A  S  V  L  S  G  R  I  A  Y  T  F  G  W  E  G  P  A  -

17623 CTGACGGGTGGACACCGGCGTCCTCGTCGGCTGCTCGTCGGCATCCACCTCGCGATGCAGGCG 17682
      --------+---------+---------+---------+---------+---------+--
       L  T  V  D  T  A  C  S  S  S  L  V  G  I  H  L  A  M  Q  A  -

17683 CTGCGGGGCGGGGTGAGTGCTCCCCTGGCGCTGGCCGTCACGGTCATGTCCGACCCG 17742
      --------+---------+---------+---------+---------+---------+--
       L
```

FIG.2-44

FIG.2-45

```
        L   R   R   G   E   C   S   L   A   L   A   G   G   V   T   V   M   S   D   P   -
17743 TACACCTTCGTCGACTTCAGCACGCAGCGCGGCTCGCCTTCCGACGGTCGCTGCAAGGCG 17802
      ---------+---------+---------+---------+---------+---------+
        Y   T   F   V   D   F   S   T   Q   R   G   L   A   S   D   G   R   C   K   A   -
17803 TTCTCCGGCGGCGGGGCCGACGGGCTTCGCGCTGTGTCGGAAGGCGTCGCCGGCGTGGTGCTGGAG 17862
      ---------+---------+---------+---------+---------+---------+
        F   S   A   R   A   D   G   F   A   L   S   E   G   V   A   A   L   V   L   E   -
17863 CCGCTTCCCGGCGCGGCCAACGGGCACCAGGTGCTGGCCGTGCTGCGGGCAGCGCG 17922
      ---------+---------+---------+---------+---------+---------+
        P   L   S   R   A   R   A   N   G   H   Q   V   L   A   V   L   R   G   S   A   -
17923 GTCAACCAGGACGGTGCCAGCAACGGTCTCGCCGCTCCCAACGGCCCGTCGCAGGAGCGG 17982
      ---------+---------+---------+---------+---------+---------+
        V   N   Q   D   G   A   S   N   G   L   A   A   P   N   G   P   S   Q   E   R   -
17983 GTGATCCGGCAGGCGCTCGCCGCTTCGGGCGTGCCGGCGGACGTCGTGGAG 18042
      ---------+---------+---------+---------+---------+---------+
        V   I   R   Q   A   L   A   A   S   G   V   P   A   A   D   V   D   V   V   E   -
```

FIG.2-46

```
18043 GCGCACGGGACGGGCACCGAGCTCGGCGACCCGATCGAGGCCGGCGCTCATCGGCGACC 18102
      ---------+---------+---------+---------+---------+---------+ -
        A  H  G  T  E  L  G  D  P  I  E  A  G  A  L  I  A  T

18103 TACGGCCAGGACCGGCGACCGGCCGCTCGGCGCTGCGGTGAAGACCAACATCGGCCAC 18162
      ---------+---------+---------+---------+---------+---------+ -
        Y  G  Q  D  R  P  L  R  L  G  S  V  K  T  N  I  G  H

18163 ACCCAGGCCGCGGGGGCCGGGGCCGTGATCAAGGTCGTGCTGGCGATGCGGCACGGG 18222
      ---------+---------+---------+---------+---------+---------+ -
        T  Q  A  A  A  G  V  I  K  V  V  L  A  M  R  H  G

18223 ATGCTGCCCCGGTCGTTGCACGCCGAGCTGTCCCCGCACATCGACTGGGAGTCGGGG 18282
      ---------+---------+---------+---------+---------+---------+ -
        M  L  P  R  S  L  H  A  D  E  L  S  P  H  I  D  W  E  S  G

18283 GCCGTGGAGGTGCTGCCGGAGGAGGTGCCGTGGCCGGCGGGGGAGCGCCCCCGGCGGGGCG 18342
      ---------+---------+---------+---------+---------+---------+ -
        A  V  E  V  L  R  E  E  V  P  W  P  A  G  E  R  P  R  R  A

18343 GGGGTGTCGTCCTTCGGCGTCAGCGGAACCAACGCGCACGTGATCGTCGAAGAGGCACCA 18402
      ---------+---------+---------+---------+---------+---------+ -
        G  V  S  F  G  V  S  G  T  N  A  H  V  I  V  E  E  A  P
```

```
18403 GCAGAGCAGGAGGCCCGCCCGCACCGAGAGCGGCGGTCCGCTGCCGTTCGTTCTGTGTCCGGCCGC 18462
      ----+----+----+----+----+----+----+----+----+----+----+----+
       A   E   Q   E   A   A   R   T   E   R   G   P   L   P   F   V   L   S   G   R

18463 AGCGAAGCCGTGGTCGCGGGCCCCAGGCCCGCGCTCGCCGAGCACCTGCGCGACACCCCG 18522
      ----+----+----+----+----+----+----+----+----+----+----+----+
       S   E   A   V   V   A   A   Q   A   R   A   L   A   E   H   L   R   D   T   P

18523 GAGCTCGGCCTGACCGACGCGGCGTGGACGCTCGCGACCGGCAGGGCGCGGTTCGACGTG 18582
      ----+----+----+----+----+----+----+----+----+----+----+----+
       E   L   G   L   T   D   A   A   W   T   L   A   T   G   R   A   R   F   D   V

18583 CGAGCCGCCGGTGCTCGGCGACGACCGCGCGGGCGTGTGCGCGGAGCTGGACGCGCTGGCC 18642
      ----+----+----+----+----+----+----+----+----+----+----+----+
       R   A   A   V   L   G   D   D   R   A   G   V   C   A   E   L   D   A   L   A

18643 GAGGGCCGCCCGTCGGCCGACGCCGTGGCCCGTCGCGCCTCCGCGCGTGACCTCCGCGCAAGCCGGTC 18702
      ----+----+----+----+----+----+----+----+----+----+----+----+
       E   G   R   P   S   A   D   A   V   A   P   V   T   S   A   P   R   K   P   V

18703 CTGGTCTTCCCCGGCCAGGGCGCGCAGTGGGTCGGCGCATGGCACGCGATCTGCTGGAATCC 18762
      ----+----+----+----+----+----+----+----+----+----+----+----+
       L   V   F   P   G   Q   G   A   Q   W   V   G   M   A   R   D   L   L   E   S
```

FIG.2-47

```
18763 TCCGAGGTGTTCGCCGAGTCGATGAGCCGGTGCGCCGAGGCGCTCTCGCCGCACACCGAC 18822
      ------+---------+---------+---------+---------+---------+---
       S  E  V  F  A  E  S  M  S  R  C  A  E  A  L  S  P  H  T  D

18823 TGGAAGTTGCTCGACGTCGTCCGCGGCGACGGCGGTCCCGACCCGCACGAGCGCGTCGAC 18882
      ------+---------+---------+---------+---------+---------+---
       W  K  L  L  D  V  V  R  G  D  G  G  P  D  P  H  E  R  V  D

18883 GTGCTCCAGCCGGTGCTCTTCTCGATCATGGTCTCGCTGGCCGAGCTGTGGCGCGCGCAC 18942
      ------+---------+---------+---------+---------+---------+---
       V  L  Q  P  V  L  F  S  I  M  V  S  L  A  E  L  W  R  A  H

18943 GGCGTGACCCCGGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCGGCGCACGTG 19002
      ------+---------+---------+---------+---------+---------+---
       G  V  T  P  A  A  V  V  G  H  S  Q  G  E  I  A  A  A  H  V

19003 GCGGGCGCGCTGTCGCTGGAAGCCGCGAAGGTGTGGCCCTGCGCAGCCAGGTGTTG 19062
      ------+---------+---------+---------+---------+---------+---
       A  G  A  L  S  L  E  A  A  A  K  V  V  A  L  R  S  Q  V  L

19063 CGCGAGCTCGACGACCAGGCGGGCATGGTGTCGGTGTCCCGCGTCCCGCGACGAGCTGGAG 19122
      ------+---------+---------+---------+---------+---------+---
```

FIG.2-48

FIG.2-49

```
        R   E   L   D   D   Q   G   G   M   V   S   V   G   A   S   R   D   E   L   E
19123 ACCGTGCTCGGCGCTGGGACGGCCGTGTCGGTGGCCGTGAACGGGCCTGGCACC 19182
      ------+---------+---------+---------+---------+---------+----

        T   V   L   A   R   W   D   G   R   V   A   A   A   V   N   G   P   G   T
19183 AGCGGTCGTTGCCGGGGCCGACCGCGCGGAGCTGGACGAGTTCTTCGCCGAGGCCGAGGGCGGG 19242
      ------+---------+---------+---------+---------+---------+----

        S   V   V   A   G   P   T   A   E   L   D   E   F   F   A   E   A   E   A   R
19243 GAGATGAAGCCGCGGCCGGATCGCCGTGCCGTACGCGCTCCCACTCCCCGGAGGTGGCCGGC 19302
      ------+---------+---------+---------+---------+---------+----

        E   M   K   P   R   R   I   A   V   R   Y   A   S   H   S   P   E   V   A   R
19303 ATCGAGGACCGGCTCGCGGGCCGAGCTGGGCACCATCACCGCCGTGCGGGGCTCGGTGCCG 19362
      ------+---------+---------+---------+---------+---------+----

        I   E   D   R   L   A   A   E   L   G   T   I   T   A   V   R   G   S   V   P
19363 CTGCACTCCACGGTGACCGGCGAGGTCATCGACACCTCCGGATGGACGCCTCCTACTGG 19422
      ------+---------+---------+---------+---------+---------+----

        L   H   S   T   V   T   G   E   V   I   D   T   S   A   M   D   A   S   Y   W
19423 TACCGCAACCTGCGCCGACCAGTGCTCTTCGAGCAGGCCGGTCGCGGTCTGGGTCGGAGCAG 19482
      ------+---------+---------+---------+---------+---------+----
```

FIG.2-50

```
        Y  R  N  L  R  R  P  V  L  F  E  Q  A  V  R  G  L  V  E  Q
19483   GGCTTCGACACCTTCGTCGAGGTGAGCCCGCACCCGGTGCTGCTGATGGCGGTCGAGGAG   19542
        ---------+---------+---------+---------+---------+---------+

        G  F  D  T  F  V  E  V  S  P  H  P  V  L  L  M  A  V  E  E
19543   ACCGCCGAGCACGCGGGCGCGGAAGTCACCTGCGTGCCGACGCTGCGCCGAGCAGAGC    19602
        ---------+---------+---------+---------+---------+---------+

        T  A  E  H  A  G  A  E  V  T  C  V  P  T  L  R  R  E  Q  S
19603   GGACCGCACGAGTTCCTGCGCAACCTGCTGCGGGCTCACGTGCACGGCGTCGGCGCCGAC   19662
        ---------+---------+---------+---------+---------+---------+

        G  P  H  E  F  L  R  N  L  L  R  A  H  V  H  G  V  G  A  D
19663   CTGCGGTCCGGGGTGGCCGGGACGGCCGGAGCTGCCCACCTACCCGTTCGAACAC       19722
        ---------+---------+---------+---------+---------+---------+

        L  R  P  A  V  A  G  G  R  P  A  E  L  P  T  Y  P  F  E  H
19723   CAGGCGCTTCTGGCCGGCCGCACCGGCCCGCCGACGTCTCGGGCGTGCGCGGC         19782
        ---------+---------+---------+---------+---------+---------+

        Q  R  E  W  P  R  P  H  R  P  A  D  V  S  A  L  G  V  R  G
        GCGGAGCACCCGCTGCTGCGCTCGCCCGGTGCTCGACGTTGCCGGGCGGTGCGGGTGTTC
```

FIG.2-51

```
19783 ------+---------+---------+---------+---------+---------+ 19842
        A   E   H   P   L   L   L   A   A   V   D   V   P   G   H   G   G   A   V   E   -
      ACCGGAAGGCTTTCCACCGACGAGCAGCCGTGGCTGGCCGAACACGTCGTGGGCGGCCGG

19843 ------+---------+---------+---------+---------+---------+ 19902
        T   G   R   L   S   T   D   E   Q   P   W   L   A   E   H   V   V   G   G   R   -
      ACGCTGGTGCCGGGCAGCGTCCTGGTCGATCTCGCGCCGCGGGTGAGGACGTCGGG

19903 ------+---------+---------+---------+---------+---------+ 19962
        T   L   V   P   G   S   V   L   V   D   L   A   L   A   A   G   E   D   V   G   -
      CTGCCGGTCCTGGAGGAACTGGTGTTGCAACGGCCGCTGGTGCTGGCCGGGGGGCG

19963 ------+---------+---------+---------+---------+---------+ 20022
        L   P   V   L   E   E   L   V   L   Q   R   P   L   V   L   A   G   A   G   A   -
      CTGCTGCGCATGTCGGTCGGCGGCCCGACGAGTCGGGGGGACGATCGACGTCCAC

20023 ------+---------+---------+---------+---------+---------+ 20082
        L   L   R   M   S   V   G   A   P   D   E   S   G   R   R   T   I   D   V   H   -
      GCCGCCGAAGACGTGGCCGACCTCGCGCCGACGGTCGCAGCACGCCACCGGGACG

20083 ------+---------+---------+---------+---------+---------+ 20142
        A   A   E   D   V   A   D   L   A   D   A   Q   W   S   Q   H   A   T   G   T   -
```

FIG.2-52

```
       .CTCGCGCAGGGCGTCGCCGCGGGTCCGAGGGATACCGAGCAGTGGCCGCCGGAGGACGCC
20143  ---------+---------+---------+---------+---------+---------+   20202
        L  A  Q  G  V  A  A  G  P  R  D  T  E  Q  W  P  P  E  D  A

       GTCCGCATCCCGCTCGACGACCACTACGACGGCCTCGCCGAGCAGGGCTACGAGTACGGA
20203  ---------+---------+---------+---------+---------+---------+   20262
        V  R  I  P  L  D  D  H  Y  D  G  L  A  E  Q  G  Y  E  Y  G

       CCGTCGTTCCAGGCCCTGCGAGCCGTGGGCGCAAGGACGACTCGGTCTACGCCGAGGTG
20263  ---------+---------+---------+---------+---------+---------+   20322
        P  S  F  Q  A  L  R  A  A  W  R  K  D  D  S  V  Y  A  E  V

       TCCATCGGCGGACGAGGAAGGTTACGCGTTCCACCCGGTGCTGCTCGACGCGCCGTGGCG
20323  ---------+---------+---------+---------+---------+---------+   20382
        S  I  A  A  D  E  E  G  Y  A  F  H  P  V  L  L  D  A  V  A

       CAGACGCTCAGCCTGGGCGCCCTCGGCGAGCCCGGGGGCGGGGAAAGCTGCCGTTCGCGGTGG
20383  ---------+---------+---------+---------+---------+---------+   20442
        Q  T  L  S  L  G  A  L  G  E  P  G  G  G  K  L  P  F  A  W

       AACACCGTGACCCTGCACGCCTCCGGTGCGGGTCGTTGGCGACGCCCGCC
20443  ---------+---------+---------+---------+---------+---------+   20502
        N  T  V  T  L  H  A  S  G  A  T  S  V  R  V  V  A  T  P  A
```

FIG.2-53

```
20503 GGGGCGGACGGCGATGGCCCTGCGGGTCACCGACCCGGCCAGGCCACCTGGTCGCCACGGTC  20562
      ---------+---------+---------+---------+---------+---------+---
            G  A  D  A  M  A  L  R  V  T  D  P  A  G  H  L  V  A  T  V  -

20563 GACTCGCTGGTCGTCCGCAGCACCGGGGGAGAAGTGGGACAGCCCGGCGAACCGGCGGTGGC  20622
      ---------+---------+---------+---------+---------+---------+---
            D  S  L  V  V  R  S  T  G  E  K  W  E  Q  P  E  P  R  G  G  -

20623 GAGGGCGAGCTGCACGCTCTCTGGACTGGGGACGGCTAGCCGGACGCCCGAGCCCGGCTCGACCGGTCGT  20682
      ---------+---------+---------+---------+---------+---------+---
            E  G  E  L  H  A  L  D  W  G  R  L  A  E  P  G  S  T  G  R  -

20683 GTGGTCGCGGCCGATGCCTCGGACCTCGAGACCGCGTCCTGCGGTCCGGTGAACCCGAACCC  20742
      ---------+---------+---------+---------+---------+---------+---
            V  V  A  A  D  A  S  D  L  D  A  V  L  R  S  G  E  P  E  P  -

20743 GACGCGGTCCTGGTCCGCTACGAACCCGAAGGCGACGACCCCGCGGCCGCCCGCCAC  20802
      ---------+---------+---------+---------+---------+---------+---
            D  A  V  L  V  R  Y  E  P  E  G  D  D  P  R  A  A  A  R  H  -

20803 GGCGTCCCTCTGGGCCGCTGGTGCGTCGAACAGGAGGAGCTGCCGGGGC  20862
      ---------+---------+---------+---------+---------+---------+---
```

```
        G  V  L  W  A  A  A  L  V  R  R  W  L  E  Q  E  E  L  P  G
20863   GCGACGCTGGTCATCGCCACGTCCGGGCGGTCCGACGACGACAGCGTTCCC           20922
        --------+---------+---------+---------+---------+---------+--

        A  T  L  V  I  A  T  S  G  A  V  T  V  S  D  D  D  S  V  P
20923   GAACCCGGCGCCGCCGGCGATGTGGGGCGTGATCCGCTGCGCAGGCCGAGTCGCCGGAC   20982
        --------+---------+---------+---------+---------+---------+--

        E  P  G  A  A  A  M  W  G  V  I  R  C  A  Q  A  E  S  P  D
20983   CGGTTCGTGCTCCTCGACACCGACGCGGAACCTGGGATGCTGCCTGCGGTTCCGGACAAC   21042
        --------+---------+---------+---------+---------+---------+--

        R  F  V  L  L  D  T  D  A  E  P  G  M  L  P  A  V  P  D  N
21043   CCGCAGCTCGGCGTTGCGGGCGACGACGTCTTCGTGCCGGGCCTCTCGCCGCTCGCACCT   21102
        --------+---------+---------+---------+---------+---------+--

        P  Q  L  A  L  R  G  D  D  V  F  V  P  R  L  S  P  L  A  P
21103   TCCGCGCTGACGCTTCCGGCAGGCACCCAACGTCTCGTGCCGGGGTGACGGGGGCGATCGAC  21162
        --------+---------+---------+---------+---------+---------+--

        S  A  L  T  L  P  A  G  T  Q  R  L  V  P  G  D  G  A  I  D
21163   TCCGTGGCCTTCGAGCCCGACACCCGACGTCGAGCAGCCGCTCCGGCGGGGGCGAGGTCCGG  21222
        --------+---------+---------+---------+---------+---------+--
```

FIG.2-54

FIG.2—55

```
            S   V   A   F   E   P   A   P   D   V   E   Q   P   L   R   A   G   E   V   R      -
      GTGGACGTGCGCGGCCACCGGAGTCAACTTCCGGCGTCCTCGCACTCGGCATGTAT
21223 ------+---------+---------+---------+---------+---------+ 21282

            V   D   V   R   A   T   G   V   N   F   R   D   V   L   L   A   L   G   M   Y      -
      CCGCAGAAGGCGGACATGGGCACCGAGGCCGCCGGTGTCGTCACGGCGGTCGGACCGGAC
21283 ------+---------+---------+---------+---------+---------+ 21342

            P   Q   K   A   D   M   G   T   E   A   A   G   V   V   T   A   V   G   P   D      -
      GTGGACGCCTTCGCGGCCGGAGACCGGGTGCTCGGCCTGTTCCAGGGAGCCTTCGCGCCG
21343 ------+---------+---------+---------+---------+---------+ 21402

            V   D   A   F   A   P   G   D   R   V   L   G   L   F   Q   G   A   F   A   P      -
      ATCGCGGTCACCGATCACCGGCTCCTGCTCGCACGAGTGCCGGACGGCTGGAGCGACGCCGAC
21403 ------+---------+---------+---------+---------+---------+ 21462

            I   A   V   T   D   H   R   L   L   A   R   V   P   D   G   W   S   D   A   D      -
      GCCGCGGGCCGTGCCCATCGCCTACACCGCGGCCATTACGGGCTGCACGATCTCGCGGGG
21463 ------+---------+---------+---------+---------+---------+ 21522

            A   A   A   V   P   I   A   Y   T   T   A   H   Y   A   L   H   D   L   A   G      -
      CTGCGCGGCGGGGTCAGTCGGTGCTCATCCACGCAGGCGGTGTCGGCATGGCGGCC
21523 ------+---------+---------+---------+---------+---------+ 21582
```

```
          L   R   A   G   Q   S   V   L   I   H   A   A   A   G   G   V   G   M   A   A   -

      GTCGCGCTGGCCCGCCGAGCGGGGGCGGAGGTGTTGGCCACCGCCGGCCCGGCCAAGCAC
21583 ---------+---------+---------+---------+---------+---------+  21642

          V   A   L   R   R   A   G   A   E   V   L   A   T   A   G   P   A   K   H   -

      GGGACGCTGCGGGCGCTCGGTCTCGACGACGAGCACATCGCTTCCTCCCGGGAGACCGGT
21643 ---------+---------+---------+---------+---------+---------+  21702

          G   T   L   R   A   L   G   L   D   D   D   E   H   I   A   S   S   R   E   T   G   -

      TTCGCCCGGAAGTTCCGGGAGCGCACCGGAGGCCGCGTGGTGTGCTCAACTCG
21703 ---------+---------+---------+---------+---------+---------+  21762

          F   A   R   K   F   R   E   R   T   G   G   R   G   V   D   V   V   L   N   S   -

      CTCACCGGGAACTGCTCGACGAGTCCGCGGATCTGCTCGCCGAGGACGGGCGTCTTCGTC
21763 ---------+---------+---------+---------+---------+---------+  21822

          L   T   G   E   L   L   D   E   S   A   D   L   L   A   E   D   G   V   F   V   -

      GAGATGGGCAAGACCGACCTGCGGGGGACTTCCGGGGCCGATACGCCCCGTTC
21823 ---------+---------+---------+---------+---------+---------+  21882

          E   M   G   K   T   D   L   R   D   A   G   D   F   R   G   R   Y   A   P   E   -
```

FIG. 2-56

FIG.2-57

```
21883  GACCTCGGGCGAGGCGGGTGACGACCGGCTCGGGGGAGATCCTGCGCGAGGTCGTCGGCCTG
       ---------+---------+---------+---------+---------+---------+  21942
           D  L  G  E  A  G  D  D  R  L  G  E  I  L  R  E  V  V  G  L  -

21943  CTGGGGCGCCGGGGGGAGCTCGACCGGCTCCCCGGTATCGGGCGTGGGAGCTGGATCCGGCCC
       ---------+---------+---------+---------+---------+---------+  22002
           L  G  A  G  E  L  D  R  L  P  V  S  A  W  E  L  G  S  A  P  -

22003  GCGGGCGGTTGCAGCACACATGAGCCGGGGGCGGCACGTCGGCAAGCTCGTGCTGACCCAGCCC
       ---------+---------+---------+---------+---------+---------+  22062
           A  A  L  Q  H  M  S  R  G  R  H  V  G  K  L  V  L  T  Q  P  -

22063  GCGGCCGGTGGACCCGGACGCGGCTGCTGATCACGGGTGGCACCGGCTCGGACGG
       ---------+---------+---------+---------+---------+---------+  22122
           A  P  V  D  P  D  G  T  V  L  I  T  G  G  T  G  T  L  G  R  -

22123  CTGCTCGCGCGGCCACCTCGTCACCGAGCACGGCGTGCGGCACCTGCTCGTGGTCAGCAGG
       ---------+---------+---------+---------+---------+---------+  22182
           L  L  A  R  H  L  V  T  E  H  G  V  R  H  L  L  L  V  S  R  -

22183  CGCGGCGCGGACGCGCGCGGGTTCCGACGAGCTGCGCGGGAGATCGAGGACTTGGGGCGCG
       ---------+---------+---------+---------+---------+---------+  22242
```

FIG.2-58

```
         R   G   A   D   A   P   G   S   D   E   L   R   A   E   I   E   D   L   G   A
         TCCGGGAGATCGGCGGCTTGCCGACACCGCCGACCGGCGCTTCGGCGCTGCTGGAC   -
22243    ---------+---------+---------+---------+---------+---------+--   22302

         S   A   E   I   A   A   C   D   T   A   D   R   D   A   L   S   A   L   L   D
         GGGCTGCCCCGGCCGCTGACCGGTGTCGTGCACGCGCGGGGTGTGCTGGCCGACGGGGCTG   -
22303    ---------+---------+---------+---------+---------+---------+--   22362

         G   L   P   R   P   L   T   G   V   V   H   A   A   G   V   L   A   D   G   L
         GTCACCTCCATCGACGAGCCGGCGGTGGAGCAGGTGCTGCGCGCCAAGGTCGACGCGGCG   -
22363    ---------+---------+---------+---------+---------+---------+--   22422

         V   T   S   I   D   E   P   A   V   E   Q   V   L   R   A   K   V   D   A   A
         TGGAACCTGCACGAGCTGACCGCGAACACCGGTCTGAGCTTCTTCGTGCTGTTCTCGTCC   -
22423    ---------+---------+---------+---------+---------+---------+--   22482

         W   N   L   H   E   L   T   A   N   T   G   L   S   F   F   V   L   F   S   S
         GCGGCGTCGGTGCTAGCCGGCCCGGGGCCGCAGGGCGTGTACGCGGCGAACGAGTCGCTC   -
22483    ---------+---------+---------+---------+---------+---------+--   22542

         A   A   S   V   L   A   G   P   G   Q   G   V   Y   A   A   N   E   S   L
         AACGCGCTGGCTGCCCTCCGGAGGACGCGGCCTTCCCGGAAGGCGCTCGGATGGGGA   -
22543    ---------+---------+---------+---------+---------+---------+--   22602
```

FIG.2-59

```
N  A  L  A  A  L  R  R  T  R  G  L  P  A  K  A  L  G  W  G  -
   CTGTGGGGCGCAGGCCAGCGAGATGACCAGCGGACTCGGCGGACCGCATCGCCGGACCGGG
22603 ---------+---------+---------+---------+---------+---------+ 22662

L  W  A  Q  A  S  E  M  T  S  G  L  G  D  R  I  A  R  T  G  -
   GTCGCCGCGCTGCCGACCGGAGCGGCGGCCTCGCACTGTTCGACAGCGCCCTGCGCCGCGC
22663 ---------+---------+---------+---------+---------+---------+ 22722

V  A  A  L  P  T  E  R  A  L  F  D  S  A  L  R  R  G  -
   GGTGAGGTCGTGTTCCCGCTGTCCATCAACCGTTCCGCGCTGCGCAGGGCCGAGTTCGTG
22723 ---------+---------+---------+---------+---------+---------+ 22782

G  E  V  V  F  P  L  S  I  N  R  S  A  L  R  R  A  E  F  V  -
   CCGGAGGTCCTGCGCGGCATGGTCAGGGCGGAAGCTGCGCGGGCCGCAGGCCGAGGGCG
22783 ---------+---------+---------+---------+---------+---------+ 22842

P  E  V  L  R  G  M  V  R  A  K  L  R  A  A  G  Q  A  E  A  -
   GCAGGGGCCGAACGTGGTCGACCGGCTCGGCGGTCCGAGTCCGACCAGGTCGCCGGG
22843 ---------+---------+---------+---------+---------+---------+ 22902

A  G  P  N  V  V  D  R  L  A  G  R  S  E  S  D  Q  V  A  G  -
   CTGGCCGAACTGGTGCGTTCACACGCGGGTCTCCGGGTACGGCTCGGCCGACCAG
22903 ---------+---------+---------+---------+---------+---------+ 22962
```

FIG.2-60

```
          L  A  E  L  V  R  S  H  A  A  A  V  S  G  Y  G  S  A  D  Q  -
22963     CTCCCCGAGGGCCAAGGCGGTTCAAGGACCTCGGTTTCGACTCGCTGGCCGCGGTGGAGCTG
          ---------+---------+---------+---------+---------+---------+  23022

          L  P  E  R  K  A  F  K  D  L  G  F  D  S  L  A  A  V  E  L  -
23023     CGCAACCGCCTCGGTACCGGCGACCGGCGTGCGGCTGCCCAGCACGTTGGTGTGTTCGACCAC
          ---------+---------+---------+---------+---------+---------+  23082

          R  N  R  L  G  T  A  T  G  V  R  L  P  S  T  L  V  F  D  H  -
23083     CCGACTCCGCTGGCGGTGGCCGGAACACCTGCGGGACAGGCTGTTCGCGGCCTCACCGGCG
          ---------+---------+---------+---------+---------+---------+  23142

          P  T  P  L  A  V  A  E  H  L  R  D  R  L  F  A  A  S  P  A  -
23143     GTGGACATCGGCGACCGGCTGGACGAGCTGGAGAAGGCGCTCGAAGCCCTGTCCGCCGAG
          ---------+---------+---------+---------+---------+---------+  23202

          V  D  I  G  D  R  L  D  E  L  E  K  A  L  E  A  L  S  A  E  -
23203     GACGGGCACGACGACGTGGGCCGGCCTGGAGTCGCTGCGCGGTGGAACAGCAGG
          ---------+---------+---------+---------+---------+---------+  23262

          D  G  H  D  D  V  G  Q  R  L  E  S  L  L  R  R  W  N  S  R  -
          CGGGCGGACGCCCCGAGCACGTCCGCGATCAGCGAGGACGACCAGTGACGACGAGCTGTTC
```

FIG.2—61

```
23263                                                                          23322
          R   A   D   A   P   S   T   S   A   I   S   E   D   A   S   D   D   E   L   F
          TCGATGCTCGACCAGCGGTTCGGGGGAGAGGACCTGTAGATGAGCGGTGACAACGGC

23323                                                                          23382
          S   M   L   D   Q   R   F   G   G   G   E   D   L   *   M   S   G   D   N   G
          ATGACCGAGGAAAAGCTCCGGCGCTACCTCAAGCGCACCGTCACCGAGCTCGACTCGGTG

23383                                                                          23442
          M   T   E   E   K   L   R   R   Y   L   K   R   T   V   T   E   L   D   S   V
          ACCGCGCGCCTGCGTGAAGTCGAGCACCGGGCCGGTGAGCCGATCGCGTCGGCATG

23443                                                                          23502
          T   A   R   L   R   E   V   E   H   R   A   G   E   P   I   A   I   V   G   M
          GCGTGCCGGTTCCCCGGCGACGTGGACTCGCCGGAGTCGTTCTGGGAGTTCGTGTCCGGC

23503                                                                          23562
          A   C   R   F   P   G   D   V   D   S   P   E   S   F   W   E   F   V   S   G
          GGCGGGGGACGCCATCGGCGGAGGCCCCCGGCCCCGAGCCGGACCGCGCTGGGAGCCGCCGACGCCG

23563                                                                          23622
          G   G   D   A   I   A   E   A   P   A   D   R   G   W   E   P   D   P   P   D   A
```

```
23623 CGGCTGGGGCGGGGATGCTCGGCGGCCGCGGGGCGGGGGCCGGGGCGACTTCGACGCGGGGCTTCTTCGGGATCTCG 23682
            ---------+---------+---------+---------+---------+---------+---------+
        R   L   G   G   M   L   A   A   A   G   D   F   D   A   G   F   F   G   I   S

23683 CCGCGCGGAGGCGCTGGCGATGGACCCGCAGCAGCGGGATCATGCTGGAGATCTCGTGGGAG 23742
            ---------+---------+---------+---------+---------+---------+
        P   R   E   A   L   A   M   D   P   Q   Q   R   I   M   L   E   I   S   W   E

23743 GCGCTGGAGCGCGCCGGCCACGATCCGGTGTCCCTGCGCGGGGCCAGCGGCGACCGGGGTGTTC 23802
            ---------+---------+---------+---------+---------+---------+
        A   L   E   R   A   G   H   D   P   V   S   L   R   G   S   A   T   G   V   F

23803 ACCGGTGTGTCGGCACCGTGGACTACGGCCCGCGACCCGGACGAGGCCCCGGACGAGGTCCTG 23862
            ---------+---------+---------+---------+---------+---------+
        T   G   V   G   T   V   D   Y   G   P   R   P   D   E   A   P   D   E   V   L

23863 GGCTACGTCGGCACCGGCACCGGCCTCCAGCGTGCCGCTCCGGGGTCGCCTACTGCCTG 23922
            ---------+---------+---------+---------+---------+---------+
        G   Y   V   G   T   G   T   A   S   S   V   A   S   G   R   V   A   Y   C   L

23923 GGCCTGGAAGGCCCGGCCGATGACCGTCGACACCGCCTGTTCCTCCGGGCTCACCGGCCCTG 23982
            ---------+---------+---------+---------+---------+---------+
        G   L   E   G   P   A   M   T   V   D   T   A   C   S   S   G   L   T   A   L
```

FIG.2-62

FIG.2-63

```
23983  CACCTGGCGATGGAGTCGCTGCGCGGGACGAGTGCGGCCTGGCGCTGGCCGGGCGGCGTG  24042
       ------+---------+---------+---------+---------+---------+---
        H   L   A   M   E   S   L   R   R   D   E   C   G   L   A   L   A   G   G   V

24043  ACGGTGATGAGCAGTCCCGGGGCCGTTCACCGAGTTCCGCAGCCAGGGCGGGCTCCGCGCC  24102
       ------+---------+---------+---------+---------+---------+---
        T   V   M   S   S   P   G   A   F   T   E   F   R   S   Q   G   G   L   A   A

24103  GACGGCCGCTGCAAGCCGTTCTCGAAGGCCGCCGACGGGTTCGGCCTGGCCGAGGGTGCC  24162
       ------+---------+---------+---------+---------+---------+---
        D   G   R   C   K   P   F   S   K   A   A   D   G   F   G   L   A   E   G   A

24163  GGGGTCCTGGTGCTGCAACGGCTGTCGCCGCGGCGGGAGGGCAGACCGGTGCTGGCC  24222
       ------+---------+---------+---------+---------+---------+---
        G   V   L   V   L   Q   R   L   S   A   A   R   R   E   G   R   P   V   L   A

24223  GTGCTGCGGGGGCTCGGGGTCAACCAGGACGGGCGCAACGGGCGCTGACCGCGCCCGAGC  24282
       ------+---------+---------+---------+---------+---------+---
        G   V   L   R   G   S   A   V   N   Q   D   G   A   S   N   G   L   T   A   P   S

24283  GGACCCGCGCAGCAGCGCGGGTCATCCGCGGGGGCTGGAGAACGCCGGTGTCCGGGGCGGGC  24342
       ------+---------+---------+---------+---------+---------+---
```

```
          G   P   A   Q   Q   R   V   I   R   R   A   L   E   N   A   G   V   R   A   G   -
          GACGTCGACTACGTGGAGGCCCACGGCACCGGGCTGGGGCGCACCCCATCGAGGTG              24402
24343     ---------+---------+---------+---------+---------+---------+

          D   V   D   Y   V   E   A   H   G   T   G   T   R   L   G   D   P   I   E   V   -
          CACGCGCTGCTCTCGACCTACGGGCGGGAACGGCGACCCGGACGATCCACTGTGGATCGGT        24462
24403     ---------+---------+---------+---------+---------+---------+

          H   A   L   L   S   T   Y   G   A   E   R   D   P   D   D   P   L   W   I   G   -
          TCGGTCAAGTCCAACATTGGCCACACCCAGGCCGCGGCCGTCGCCGGGGTGATGAAG            24522
24463     ---------+---------+---------+---------+---------+---------+

          S   V   K   S   N   I   G   H   T   Q   A   A   A   G   V   A   G   V   M   K   -
          GCGGTGCTGGCGCTGCGGCACGGGCGAGATGCCGCGGACTTCGACGAGCCCTCG               24582
24523     ---------+---------+---------+---------+---------+---------+

          A   V   L   A   L   R   H   G   E   M   P   R   T   L   H   F   D   E   P   S   -
          CCGCAGATCGAGTGGGACCTGGGCGCGGTGTCGGTGTGTCGCAGGCGGTCGTGGCCC            24642
24583     ---------+---------+---------+---------+---------+---------+

          P   Q   I   E   W   D   L   G   A   V   S   V   V   S   Q   A   R   S   W   P   -
          GCCGGCGGAGAGGCCCCGCAGGGGCGTCTCCTCGTTCGGCATCAGCGGCACCAACGCG           24702
24643     ---------+---------+---------+---------+---------+---------+
```

FIG.2-64

SUBSTITUTE SHEET

FIG.2-65

```
        A  G  E  R  P  R  R  A  G  V  S  S  F  G  I  S  G  T  N  A  -
24703   CACGTCATCGTCGAAGAGAGGCGCCCGAGGCCGGCCGAGAGCCGGCACCCGACTCGGGT   24762
        ----+----+----+----+----+----+----+----+----+----+----+----+--

        H  V  I  V  E  E  A  P  E  A  D  E  P  E  P  A  P  D  S  G  -
24763   CCGGTCCGCTGGTGTTGTCCGGCCGCGGCGACGAGCAGGCGGGGGCGCAGGGGGACGG   24822
        ----+----+----+----+----+----+----+----+----+----+----+----+--

        P  V  P  L  V  L  S  G  R  D  E  Q  A  M  R  A  Q  A  G  R  -
24823   CTGGCAGACCACCTCGCCCGGCCCGGAGCCCGGGAACTCGTTGCGCGACACCGGTTCACGCTG   24882
        ----+----+----+----+----+----+----+----+----+----+----+----+--

        L  A  D  H  L  A  R  E  P  P  R  N  S  L  R  D  T  G  F  T  L  -
24883   GCCACCCGCCGCAGCCGGCGGTGGGAGCACCCGCGCGGGTCGGCGACCGCGACGACGCC   24942
        ----+----+----+----+----+----+----+----+----+----+----+----+--

        A  T  R  R  S  A  W  E  H  R  A  V  V  V  G  D  R  D  D  A  -
24943   CTCGCCGGGCTGCGCGGGCTCGGTGGCGCGGCCCGACGGCCGCATCGCCGACGGCCACCGGGGCAG   25002
        ----+----+----+----+----+----+----+----+----+----+----+----+--

        L  A  G  L  R  A  V  A  D  G  R  I  A  D  R  T  A  T  G  Q  -
25003   GCCCGAACTCGCCGGCGTCGCGGATGGTGTTCCCCGGCCCAGGGGCGCCAGTGGCAGGGGG   25062
        ----+----+----+----+----+----+----+----+----+----+----+----+--
```

```
      A  R  T  R  R  G  V  A  M  V  F  P  G  Q  G  A  Q  W  Q  G
      ATGGCCCGGACCTGCTGCGGGGAGTCGCAGGTATTCGCCGGACTCGCGGACTGCGAG
25063 ---------+---------+---------+---------+---------+--------- 25122

      M  A  R  D  L  L  R  E  S  Q  V  E  A  D  S  I  R  D  C  E
      CGGGCGCTGGCCCCGCACGTCGACTGGTCGCTGACCGACCTGCTCAGCGGCGCGACCG
25123 ---------+---------+---------+---------+---------+--------- 25182

      R  A  L  A  P  H  V  D  W  S  L  T  D  L  L  S  G  A  R  P
      CTGGACCGGGTCGACGTCGTCCAGCCCGCCCGGCTCTTCGCCGTCATGGTGTCGCTGGCGGCG
25183 ---------+---------+---------+---------+---------+--------- 25242

      L  D  R  V  D  V  V  Q  P  A  L  F  A  V  M  V  S  L  A  A
      CTGTGGCGCTCCCACGGCGTCGAGCCCGCCGCGGTCGTCGGCCACTCGCAGGGCGAGATC
25243 ---------+---------+---------+---------+---------+--------- 25302

      L  W  R  S  H  G  V  E  P  A  A  V  V  G  H  S  Q  G  E  I
      GCCGCCGGCGCACGTCGCCGCGGCGCTCACCCTGGAGGACGCCGCCAAGCTCGTCGGGTC
25303 ---------+---------+---------+---------+---------+--------- 25362

      A  A  H  V  A  G  A  L  T  L  E  D  A  A  K  L  V  A  V
```

FIG.2-66
```

FIG.2–67

```
25363                                                              25422
     CGGAGCCGGGTCCTGGCCCGGCTCGGGCCGGCCAGGGCGGCCATGGCGTCGTTCGGGCTGGGC
     ---------+---------+---------+---------+---------+---------+
       R  S  R  V  L  A  R  L  G  G  Q  G  G  M  A  S  F  G  L  G  -

25423                                                              25482
     ACCGAGCAGGCGGCCGAACGGATCGGGCGGCTTCGCGCGCTCTCCATCGCCTCGGTC
     ---------+---------+---------+---------+---------+---------+
       T  E  Q  A  A  E  R  I  G  R  F  A  G  A  L  S  I  A  S  V  -

25483                                                              25542
     AACGGCCCCGGTCGTCGTCGCGGGGGAGAGCGGGCCGCTGGACGAGCTGATCGCC
     ---------+---------+---------+---------+---------+---------+
       N  G  P  R  S  V  V  V  A  G  E  S  G  P  L  D  E  L  I  A  -

25543                                                              25602
     GAGTGCGAGGCCGAAGGCATAACGGCGCCGCATCCCCGTCGACTACGCCTCCCACTCA
     ---------+---------+---------+---------+---------+---------+
       E  C  E  A  E  G  I  T  A  R  R  I  P  V  D  Y  A  S  H  S  -

25603                                                              25662
     CCGCAGGTTGGAGTCGCTGCGCGAGGAGCTGCTGACCGAGCTGGCGGCATCTCCCCGGTG
     ---------+---------+---------+---------+---------+---------+
       P  Q  V  E  S  L  R  E  E  L  L  T  E  L  A  G  I  S  P  V  -

25663                                                              25722
     TCGGGCGGACGTGGCCGGCTCTACTCGACCACGACCGGGCCATCGACACCGCCACGATG
     ---------+---------+---------+---------+---------+---------+
```

FIG.2-68

```
        S  A  D  V  A  L  Y  S  T  T  T  G  Q  P  I  D  T  A  T  M
25723  GACACCGGCCTACTGGTACGCGAACCTGCGGCGGAGCAGGTCCGCTTCCAGGACGCGGACGCGG  25782
       -----+---------+---------+---------+---------+---------+---------+

        D  T  A  Y  W  Y  A  N  L  R  E  Q  V  R  F  Q  D  A  T  R
25783  CAGCTCGCCGAGGGCGGGGTTCGACGCGGTTCGTCGAGGTCAGCCCGCATCCGGTGCTGACC  25842
       -----+---------+---------+---------+---------+---------+---------+

        Q  L  A  E  A  G  F  D  A  F  V  E  V  S  P  H  P  V  L  T
25843  GTCGGCATCGAGGCCACGCTGGACTCCGGCTCCCGCGCCGGGCCTGCGTCGTG  25902
       -----+---------+---------+---------+---------+---------+---------+

        V  G  I  E  A  T  L  D  S  A  L  P  A  D  A  G  A  C  V  V
25903  GGCACCCTGCGCGGGACCGGCGGCCTGGCCGACTTCCACACCGGCTCGGCGGAGGCCG  25962
       -----+---------+---------+---------+---------+---------+---------+

        G  T  L  R  R  D  R  G  G  L  A  D  F  H  T  A  L  G  E  A
25963  TACGCGCAGGGCGTGGAGGTCGACTGGAGTCGACTGGAGCCCCGCCTTCGCCGACGCGCGGTCGGAG  26022
       -----+---------+---------+---------+---------+---------+---------+

        Y  A  Q  G  V  E  V  D  W  S  P  A  F  A  D  A  R  P  V  E
26023  CTGCCCGTCTACCCGTTCCAGCGGCCAGCGGTACTGGCTGCCCATCCCCACCGGCGGGGGGC  26082
       -----+---------+---------+---------+---------+---------+---------+
```

```
      L P V Y P F Q R Q R Y W L P I P T G G R -
26083 GCACGGGACGAGGAGGACGACGACTGGCGCGCTACCAGTCGTATGGGGGAAGCCGAGTGGGAG
      ---------+---------+---------+---------+---------+---------+ 26142

      A R D E D D D W R Y Q V V W R E A E W E -
26143 AGCGCTTCGCTGGCCGGACGCGTCGTGCTGTGGTGACCGGACCGGGCGTCCGTCCGAGTTG
      ---------+---------+---------+---------+---------+---------+ 26202

      S A S L A G R V L L V T G P G V P S E L -
26203 TCGGACGCGCCATCCGAAGTGGACTGGAGCAGAGCGGGTGCGACGGTCCTGACCTGCGACGTG
      ---------+---------+---------+---------+---------+---------+ 26262

      S D A I R S G L E Q S G A T V L T C D V -
26263 GAATCCCGTTCGACCATCGGCACCGCACTGGAGGCCGCCGACACCGACGCTCTGTCCACT
      ---------+---------+---------+---------+---------+---------+ 26322

      E S R S T I G T A L E A A D T D A L S T -
26323 GTGGTGTCGCTGCTGTCCCGCGACGGCGAGGCCGTCGATCCGTCGCTGGACGCGCTCGCC
      ---------+---------+---------+---------+---------+---------+ 26382

      V V S L L S R D G E A V D P S L D A L A -
26383 CTGGTCCAGGCCCTCGGACGGGTCGAAGCACCGCTGTGGTGCTGACCCGCAAC
      ---------+---------+---------+---------+---------+---------+ 26442
```

FIG.2—69

FIG.2-70

```
      L  V  Q  A  L  G  A  A  G  V  E  A  P  L  W  V  L  T  R  N  -
      GCCGTGCAGGTGGCCGACGGGCGAACTGGTCGATCCGGCCGCAGGCCATGGTGGGCGGTCTC
26443 ---------+---------+---------+---------+---------+---------+    26502

      A  V  Q  V  A  D  G  E  L  V  D  P  A  Q  A  M  V  G  G  L  -
      GGCCGCGTGGTCGGCATCGAGCAGCCGGGGCGCTGGGGCGGTCTGGTGGACCTGGTCGGAC
26503 ---------+---------+---------+---------+---------+---------+    26562

      G  R  V  V  G  I  E  Q  P  G  R  W  G  G  L  V  D  L  V  D  -
      GCCGATGCCGCCGTCGATCCGGTCGCTGGCCGTGCTGGCCGGTGCGGGGCGGGGCGAGGAG
26563 ---------+---------+---------+---------+---------+---------+    26622

      A  D  A  A  S  I  R  S  L  A  A  V  L  A  D  P  R  G  E  E  -
      CAGGTCGCGATCCGGGGGGACGGGATCAAGGTGGCCGAGGCTCGTGCCCGCCCCCGC
26623 ---------+---------+---------+---------+---------+---------+    26682

      Q  V  A  I  R  A  D  G  I  K  V  A  R  L  V  P  A  P  A  R  -
      GCCGCACGGCACCCGCTGGAGCCCTCGCGGTCACCGGGGCGCACCGGAGGG
26683 ---------+---------+---------+---------+---------+---------+    26742

      A  A  R  T  R  W  S  P  R  G  T  V  L  V  T  G  G  T  G  G  -
      ATCGGCGGCACGTCGCCCGCTGGCCCGCTCGGGCGCCGAGCACCTGGTGCTGCTG
```

FIG.2-71

```
26743 ----+----+----+----+----+----+----+----+----+----+----+----+ 26802
        GGCAGGCGCGGTGCCGACGCACCCGGGCGCGTCCGAGCTGAGGGAGGAGCTGACCGCGCTC
         I  G  A  H  V  A  R  W  L  A  R  S  G  A  E  H  L  V  L  L

26803 ----+----+----+----+----+----+----+----+----+----+----+----+ 26862
        GGCACGGGCGTGACCATCGCCGCCTGCGACGTCGCCGACGTCGCCGGGCTCGAAGCGGTG
         G  R  R  G  A  D  A  P  G  A  S  E  L  R  E  E  L  T  A  L

26863 ----+----+----+----+----+----+----+----+----+----+----+----+ 26922
        CTCGCCGCGGGAGCGGCGGAGGGACGGCACGGTCAGCGCCGTGATGCACGCGGCGGGGTT
         G  T  G  V  T  I  A  A  C  D  V  A  D  R  A  R  L  E  A  V

26923 ----+----+----+----+----+----+----+----+----+----+----+----+ 26982
        TCCACGTCCACGCCCCTCGACGACCTCACCGAAGCCGAGTTCACCGAGATCGCCGACGTG
         L  A  A  E  R  A  E  G  R  T  V  S  A  V  M  H  A  A  G  V

26983 ----+----+----+----+----+----+----+----+----+----+----+----+ 27042
        AAGGTGCGGGCACCGTCAACCTGGACGAGCTCTGCCCGGACCTCGACGCGTTCGTGTTG
         S  T  S  T  P  L  D  D  L  T  E  A  E  F  T  E  I  A  D  V

27043 ----+----+----+----+----+----+----+----+----+----+----+----+ 27102
         K  V  R  G  T  V  N  L  D  E  L  C  P  D  L  D  A  F  V  L
```

```
27103 TTCTCCTCCAACGCGGGCGGTGTGGGCAGTCCGGGGGCTCGCCTCCTACGCGGGGCCAAC 27162
      ----------+---------+---------+---------+---------+---------+ -
       F   S   S   N   A   G   V   W   G   S   P   G   L   A   S   Y   A   A   A   N

27163 GCCTTCCTCGACGGCTTCGCGCGGCGCCGGAGCGAGGGCGCCGGTGACGTCCATC 27222
      ----------+---------+---------+---------+---------+----- -
       A   F   L   D   G   F   A   R   R   R   S   E   G   A   P   V   T   S   I

27223 GCCTGGGGGCTCTGGGCCGGGCAGAACATGGCCGGGGACGAGGGCGGAGTACCTGCGC 27282
      ----------+---------+---------+---------+---------+------- -
       A   W   G   L   W   A   G   Q   N   M   A   G   D   E   G   G   E   Y   L   R

27283 AGCCAGGGCCTGCGGGCCATGGACCCGGATCGGGCCGTCGAGGAACTGCACATCACCCTC 27342
      ----------+---------+---------+---------+---------+---------+ -
       S   Q   G   L   R   A   M   D   P   D   R   A   V   E   E   L   H   I   T   L

27343 GACCACGGTCAGACGTCCGTGTCGGTGGACATGGATCGCAGGCGGTTCGTGGAGCTG 27402
      ----------+---------+---------+---------+---------+----- -
       D   H   G   Q   T   S   V   S   V   V   D   M   D   R   R   R   F   V   E   L

27403 TTCACCGGCCCGGCACCGGCCTGTTCGACGAGATCGCCGGTGCCCGGCGGGAAGCC 27462
      ----------+---------+---------+---------+---------+------ -
       F   T   A   A   R   H   R   P   L   F   D   E   I   A   G   A   R   A   E   A
```

FIG.2—72

```
27463  CGGCAGAGCGAGGAGGGCCCGGCGCTCGCCCAGCGGCTCGCGGCTGTGTCGAGGCCGAG  27522
       ------+---------+---------+---------+---------+---------+-
        R  Q  S  E  E  G  P  A  L  A  Q  R  L  A  A  L  S  T  A  E

27523  AGGCGGGAGCACCTCGCCCACCTGATCCGCGCCGAGGTCGCGGTGCTCGGCCACGGC  27582
       ------+---------+---------+---------+---------+---------+-
        R  R  E  H  L  A  H  L  I  R  A  E  V  A  A  V  L  G  H  G

27583  GACGACGCGGCGATCGACCGCGACCGCGCCTTCCGCGACCTCGGCTTCGACTCCATGACC  27642
       ------+---------+---------+---------+---------+---------+-
        D  D  A  A  I  D  R  D  R  A  F  R  D  L  G  F  D  S  M  T

27643  GCCGTCGACCTGCGCGAACCGGCTCGCGCGGTGACCGGGGTGCGCGGGAAGCCGGACGGTG  27702
       ------+---------+---------+---------+---------+---------+-
        A  V  D  L  R  N  R  L  A  A  V  T  G  V  R  E  A  A  T  V

27703  GTCTTCGACCACCCGACCATCACCCGGCTCGCCGACCACTACCTGGAGCGGCTCGTCGGC  27762
       ------+---------+---------+---------+---------+---------+-
        V  F  D  H  P  T  I  T  R  L  A  D  H  Y  L  E  R  L  V  G

27763  GCAGCAGAGGCGGAGCAAGCCCCGGCGCTCGTGCGCGAGGTGCCGAAGGATGCCGACGAC  27822
       ------+---------+---------+---------+---------+---------+-
        A  A  E  A  E  Q  A  P  A  L  V  R  E  V  P  K  D  A  D  D
```

FIG.2-73

```
27823 CCGATCGGCGATCGTCGGCATGGCCTGCCGCTTCCCCGGCGGCGTGCACAACCCCGGTGAG 27882
      -----+---------+---------+---------+---------+---------+
         P  I  A  I  V  G  M  A  C  R  F  P  G  G  V  H  N  P  G  E

27883 CTGTGGGAGTTCATCGTCGGCCGCGGGAGACGCCGTGACGGAGATGCCCACCGACCGCGGC 27942
      -----+---------+---------+---------+---------+---------+
         L  W  E  F  I  V  G  R  G  D  A  V  T  E  M  P  T  D  R  G

27943 TGGGACCTCGACGCGCTGTTCGACCCCGACCCCGACCCGCAGCGGCCACGGAACCAGCTACTCGCGA 28002
      -----+---------+---------+---------+---------+---------+
         W  D  L  D  A  L  F  D  P  D  P  Q  R  H  G  T  S  Y  S  R

28003 CACGGCGCGGTTCCTCGACGGGCCGCCGACTTCGACGCGGCCGTTCTTCGGGATCTCGCCG 28062
      -----+---------+---------+---------+---------+---------+
         H  G  A  F  L  D  G  A  A  D  F  D  A  A  F  F  G  I  S  P

28063 CGCGAGGCGCTGGCGATGGACCCGCAGCAGCGGCCCAGGTCCTGGAAACGACGTGGGAGCTG 28122
      -----+---------+---------+---------+---------+---------+
         R  E  A  L  A  M  D  P  Q  Q  R  Q  V  L  E  T  T  W  E  L

28123 TTCGAGAACGCCGGCATCGACCCGCACTCGCTGCTGGGGGCAGCGACACCGGCGTCTTCCTC 28182
      -----+---------+---------+---------+---------+---------+
```

FIG.2-74

```
          F   E   N   A   G   I   D   P   H   S   L   R   G   S   D   T   G   V   F   L
28183   GGCGGCCGCGTACCAGGGCTACGGCCCAGGACGGCGGTGGTGCCCGAGGACAGCGAGGGCTAC   28242
        ---------+---------+---------+---------+---------+---------+

          G   A   A   Y   Q   G   Y   G   Q   D   A   V   V   P   E   D   S   E   G   Y
28243   CTGCTCACCGGCAACTCCTCCGCCGTGGTGTCCGGCCGGGTCGCCGGGTCGCCTACGTGCTGGGGCTG   28302
        ---------+---------+---------+---------+---------+---------+

          L   L   T   G   N   S   S   A   V   V   S   G   R   V   A   Y   V   L   G   L
28303   GAAGGCCCCGCGGTCACGGTGGACACGGCGGCGTGTTCGTCGTTCGTTGGTGGCCTTGCATTCG   28362
        ---------+---------+---------+---------+---------+---------+

          E   G   P   A   V   T   V   D   T   A   C   S   S   S   L   V   A   L   H   S
28363   GCGTGTGGGGTCGTTGCGTTGCGTGACGGTGACTGCGGTCTTGCGGGTGGCCGGGTGGTGTGTCGGTG   28422
        ---------+---------+---------+---------+---------+---------+

          A   C   G   S   L   R   D   G   D   C   G   L   A   V   A   G   G   V   S   V
28423   ATGGGGCCCGGAGGGTGTTCACCGAGTTCTCCCGCCAGGGCGGCTTGGCCGTGGACGGG   28482
        ---------+---------+---------+---------+---------+---------+

          M   A   G   P   E   V   F   T   E   F   S   R   Q   G   G   L   A   V   D   G
28483   CGCTGCAAGGGGTTCTCCGCGGAGGCCGGGTTCGCGGAGGGCGTCGCGGTG   28542
        ---------+---------+---------+---------+---------+---------+
```

FIG.2-75

FIG.2-76

```
          R  C  K  A  F  S  A  E  A  D  G  F  G  F  A  E  G  V  A  V
          GTCCTGCTCCAGGCGGTTGTCCGACGCCCGCAGGGCGGGGGTCGCGCCAGGTGCTCGGGGTC
28543     ------+---------+---------+---------+---------+---------+---   28602

          V  L  L  Q  R  L  S  D  A  R  R  A  G  R  Q  V  L  G  V  V
          GCGGGCTCGGGCGATCAACCAGGACGGCGCGAGCAACGGTCTCGCGGCCGAGCGGCCGTC
28603     ------+---------+---------+---------+---------+---------+---   28662

          A  G  S  A  I  N  Q  D  G  A  S  N  G  L  A  A  P  S  G  V
          GCCCAGCAGCGGCGTGATCCGCAAGGCGTGGGGCGCGTGCGGGGGATCACCGGCGGGGATGTG
28663     ------+---------+---------+---------+---------+---------+---   28722

          A  Q  R  V  I  R  K  A  W  A  R  R  A  G  I  T  G  A  D  V
          GCCGTGGGTGTGGAGGCGCATGGGGACCGGGTACGGGGCGGATCCGGTGTGGAGGCGTCGGCG
28723     ------+---------+---------+---------+---------+---------+---   28782

          A  V  V  E  A  H  G  T  G  T  R  L  G  D  P  V  E  A  S  A
          TTGCTGGCTACTTACGGCAAGTCCGCGGGGGTCGTCGGGCCCGTGCTGCTGGGTTCGGTG
28783     ------+---------+---------+---------+---------+---------+---   28842

          L  L  A  T  Y  G  K  S  R  G  S  S  G  P  V  L  L  G  S  V
```

FIG.2-77

```
28843 AAGTCGAACATCGGTCACGCGCCAGGCGGCCGCGGGGTGTCGCGGGCGTGATCAAGGTGGTC 28902
      ---------+---------+---------+---------+---------+---------+---
           K  S  N  I  G  H  A  Q  A  A  A  G  V  A  G  V  I  K  V  V

28903 CTGGGGTTGAACCGCGGCCTGGTGCCGCCGATGCTCTGCCGCGGCGAGCCGGTCGCCGCTG 28962
      ---------+---------+---------+---------+---------+---------+---
        L  G  L  N  R  G  L  V  P  P  M  L  C  R  G  E  R  S  P  L

28963 ATCGAATGGTCCTCGGGTGGTGTGGAACTTGCCGAGGCCGTGAGCCCGTGGCCTCCGGCC 29022
      ---------+---------+---------+---------+---------+---------+---
        I  E  W  S  S  G  G  V  E  L  A  E  A  V  S  P  W  P  P  A

29023 GCGGACGGGGTGCGCCGGGCGGGGGTGTCGGCGTTCGGGGGTGAGCGGGACGAACGCGCAC 29082
      ---------+---------+---------+---------+---------+---------+---
        A  D  G  V  R  R  A  G  V  S  A  F  G  V  S  G  T  N  A  H

29083 GTGATCATCGCCGAGCCCCCGAGCCCGCTGCCGGAACCCGGTGGGCGTG 29142
      ---------+---------+---------+---------+---------+---------+---
        V  I  I  A  E  P  P  E  P  P  L  P  E  P  G  P  V  G  V

29143 CTGGCCGCTGCGAACTCGGTGCCCGTACTGCTGTCGGCCAGGACCGGAGACCGGTTGGCA 29202
      ---------+---------+---------+---------+---------+---------+---
```

```
                    L  A  A  A  N  S  V  P  V  L  L  S  A  R  T  E  T  A  L  A   -
29203  GCGCAGGGCGGGCTCCTGGAGTCCGCAGTGGACGACTCGGTTCCGTTGACGGCATTGGCT   29262
       ---------+---------+---------+---------+---------+---------+

                    A  Q  A  R  L  L  E  S  A  V  D  D  D  S  V  P  L  T  A  L  A   -
29263  TCCGCGCTGGCCACCGGACGCGCCCACCTGCCGCGTCGTGCGGCGTTGCTGGCAGGCGAC   29322
       ---------+---------+---------+---------+---------+---------+

                    S  A  L  A  T  G  R  A  H  L  P  R  R  A  A  L  L  A  G  D   -
29323  CACGAACAGCTCCGCGGCAGTTGCGAGCGGTCGCCGAGGGCGTTGCGCGTCCCGGTGCC   29382
       ---------+---------+---------+---------+---------+---------+

                    H  E  Q  L  R  G  Q  L  R  A  V  A  E  G  V  A  A  P  G  A   -
29383  ACCACCGGAACCGCCTCCGCGGGCGGCGTTGGTTTCGTCTTCCCAGGTCAGGGTGCTCAG   29442
       ---------+---------+---------+---------+---------+---------+

                    T  T  G  T  A  S  A  G  G  V  V  F  V  F  P  G  Q  G  A  Q   -
29443  TGGGAGGGCATGGCCCGGGGCCTTGCTCTCGGTCCCCGTCTTCGCCCGAGTCGCATCGGCCGAG   29502
       ---------+---------+---------+---------+---------+---------+

                    W  E  G  M  A  R  G  L  L  S  V  P  V  F  A  E  S  I  A  E   -
29503  TGCGATGCGGTGTTGTCGGAGGTGGCGGGTTCTCGGCCTCCGAAGTGCTGGAGCAGCGT   29562
       ---------+---------+---------+---------+---------+---------+
```

FIG.2-78

```
            C  D  A  V  L  S  E  V  A  G  F  S  A  S  E  V  L  E  Q  R  -
29563  CCGGACGCGCCGTCGCTGGAGCGGTCGACGTCGTACAGCCGGTGTTCTCCGTGATG
       --------+---------+---------+---------+---------+---------+  29622

            P  D  A  P  S  L  E  R  V  D  V  V  Q  P  V  L  F  S  V  M  -
29623  GTGTCGCTGGGCGGGGCTGTGGGGCTTGCGGAGTCAGCCCCTCGGCCGTCATCGGCCAT
       --------+---------+---------+---------+---------+---------+  29682

            V  S  L  A  R  L  W  G  A  C  G  V  S  P  S  A  V  I  G  H  -
29683  TCGCAGGGGCGAGATCGCCCCGGCCCGGGGGTGTTGCCGGGGTGTTGTCGCTGGAGGACGGGCGTG
       --------+---------+---------+---------+---------+---------+  29742

            S  Q  E  I  A  A  A  V  V  A  G  V  L  S  L  E  D  G  V  -
29743  CGGGTCGTGTGGCCCCTGCGCGCGAAGGGCGTTGCGTGCTGGGCGGGCAAGGGCGGCATGGTC
       --------+---------+---------+---------+---------+---------+  29802

            R  V  V  A  L  R  A  K  A  L  R  A  L  A  G  K  G  G  M  V  -
29803  TCGTTGGCGCTCCGGTGAACGCGCCCGCGCTGATCGCACCGTGGGAGGACCGGATC
       --------+---------+---------+---------+---------+---------+  29862

            S  L  A  A  P  G  E  R  A  R  A  L  I  A  P  W  E  D  R  I  -
29863  TCCGTCGCGGCGGTCAACTCCCCGTCCTCGGTGGTCTCCGGGATCCGGAGGCGCTG
       --------+---------+---------+---------+---------+---------+  29922
```

FIG.2-79

FIG.2—80

```
          S V A A V N S P S S V V V S G D P E A L
29923  GCCGAACTCGTGCGCACGTTGGCGAGGACGAGGGGCGTGCCGCGCCAAGACGCTCCCGGTGGAC  29982
       -----+----+----+----+----+----+----+----+----+----+----+----+-

          A E L V A R C E D E G V R A K T L P V D
29983  TACGCCTCGCACTCCCGCCACGTCGAGGAGATCCGGCGAGACGATCCTCGCCGACCTCGAC  30042
       -----+----+----+----+----+----+----+----+----+----+----+----+-

          Y A S H S R H V E E I R E T I L A D L D
30043  GGCATCTCCGCGCGGCGTGCCGCGCCATCCCGCTCTACTCCACGCTGCACGGGCGAACGGGCGC  30102
       -----+----+----+----+----+----+----+----+----+----+----+----+-

          G I S A R R A A I P L Y S T L H G E R R
30103  GACGGCGGCCGACATGGGTCCGGTCCGGTACTGGTACGACAACCTGCGCTCCCAGGTGCGCTTC  30162
       -----+----+----+----+----+----+----+----+----+----+----+----+-

          D G A D M G P R Y W Y D N L R S Q V R F
30163  GACGAGGCGGTCTCGGCCGTCGCGCCGACGGTCACGCGCCACCTTCGTCGAGATGAGCCCG  30222
       -----+----+----+----+----+----+----+----+----+----+----+----+-

          D E A V S A A V A D G H A T F V E M S P
       CACCCGGTGCTCACCGGGGTGCAGGAGATCGCCGCGGTGGCCATCGGGGTCG
```

```
30223 ---------+---------+---------+---------+---------+---------+ 30282
        H  P  V  L  T   A  A  V  Q  E  I   A  A  D  A  V  A  I  G  S  -
      CTGCACCGCGACACCGCGGAGGAGCACCTGATCGCGCCGAGTCGCGCCGGGCGCACGTGCAC
      ---------+---------+---------+---------+---------+---------+
        L  H  R  D  T   A  E  E  H  L  I   A  E  L  A  R  A  H  V  H  -

30283 ---------+---------+---------+---------+---------+---------+ 30342
      GGCGTGGCCGTGGACTGGCGGAACGTCTTCCCGGCGGCACCTCCGGTGGCGCTGCCCAAC
      ---------+---------+---------+---------+---------+---------+
        G  V  A  V  D  W  R  N  V  F  P  A  A  P  P  V  A  L  P  N  -

30343 ---------+---------+---------+---------+---------+---------+ 30402
      TACCCGTTCGAGCCCCAGCGGTACTGGCTCGCGCCGGAGGTGTCCGACCAGCTCGCCGAC
      ---------+---------+---------+---------+---------+---------+
        Y  P  F  E  P  Q  R  Y  W  L  A  P  E  V  S  D  Q  L  A  D  -

30403 ---------+---------+---------+---------+---------+---------+ 30462
      AGCCGCTACCGCGTCGACTGGCGGCCGCTGGCCACCACGCCGGTGGACCTGGAAGGCGGC
      ---------+---------+---------+---------+---------+---------+
        S  R  Y  R  V  D  W  R  P  L  A  T  T  P  V  D  L  E  G  G  -

30463 ---------+---------+---------+---------+---------+---------+ 30522
      TTCCTGGTCCACGGGTCCGCACCGGAGTCGCTGACCAGCGCCGTCGAGAAGGCCGGAGGC
      ---------+---------+---------+---------+---------+---------+
        F  L  V  H  G  S  A  P  E  S  L  T  S  A  V  E  K  A  G  G  -

30523 ---------+---------+---------+---------+---------+---------+ 30582
```

FIG.2-81

FIG.2-82

```
30583 CGCGTCGGTGCCGGTCGCCTCGGCCGGTCGCCTCGGCCGGCGAAGCCTCGGCGGCCCTGCGGGAGGTGCCG 30642
      ----+----+----+----+----+----+----+
       R  V  V  P  V  A  S  A  D  R  E  A  S  A  A  L  R  E  V  P

30643 GGCGAGGTCGCCGGCGGTGCTCTCGGTCCACACCGGCGCCAACGCACCTCGCCCTGCAC 30702
      ----+----+----+----+----+----+
       G  E  V  A  G  V  L  S  V  H  T  G  A  A  T  H  L  A  L  H

30703 CAGTCGCTGGGTGAGGCCGGCGTGCGGGCCCGCTCTGCGCTGGTCACCAGCCGAGCCGGTC 30762
      ----+----+----+----+----+----+
       Q  S  L  G  E  A  G  V  R  A  P  L  W  L  V  T  S  R  A  V

30763 GCGCTCGGGGAGTCCGAGCCGGTCGATCCCGAGCAGGCGATGGTGTGGGGTCTCGGGCGC 30822
      ----+----+----+----+----+----+
       A  L  G  E  S  E  P  V  D  P  E  Q  A  M  V  W  G  L  G  R

30823 GTCATGGGCCTGGAGACCCCGGAACGGTGGGGCGGTCTGGTGGACCTGCCCGCCGAACCC 30882
      ----+----+----+----+----+----+
       V  M  G  L  E  T  P  E  R  W  G  G  L  V  D  L  P  A  E  P

30883 GCGCCGGGGACGGGCGAGGCGGTTCGTCGGCCGCGCCTGCCTCGGCGACGGCCACGAGGACCAG 30942
      ----+----+----+----+----+----+
       A  P  G  D  G  E  A  F  V  A  C  L  G  A  D  G  H  E  D  Q
```

```
30943  GTCGCGATCCGTGACCACGCCCGCGCTACGGGCCGCGCCCTCGTCCGCGCCCCGCTGGGCACC  31002
       ---------+---------+---------+---------+---------+---------+
        V  A  I  R  D  H  A  R  Y  G  R  R  R  L  V  R  A  P  L  G  T  -

31003  CGCGAGTCGAGCTGGGAGCCGGGGGCACGGGCGCTGGTCACCGGGCGGTGCGCTC  31062
       ---------+---------+---------+---------+---------+---------+
        R  E  S  W  E  P  A  G  T  A  L  V  T  G  G  T  G  A  L  -

31063  GGCGGCCACGTCGCCCGCCACCTCGCCCAGGTGCGGGGGTGGAGGACCTGGTGCTGGTCAGC  31122
       ---------+---------+---------+---------+---------+---------+
        G  G  H  V  A  R  H  L  A  R  C  G  V  E  D  L  V  L  V  S  -

31123  AGGCGCGGCGTCGACGCTCCCGCGCGGAAGCCGAACTGGTCGCCCTCGGC  31182
       ---------+---------+---------+---------+---------+---------+
        R  R  G  V  D  A  P  G  A  A  E  L  E  A  E  L  V  A  L  G  -

31183  GCGAAGACGACCATCACCGCCTGCGACGTGGCCGACCGCGAGCAGCTCTCCAAGCTGCTG  31242
       ---------+---------+---------+---------+---------+---------+
        A  K  T  T  I  T  A  C  D  V  A  D  R  E  Q  L  S  K  L  L  -

31243  GAAGAACTGCGCGGGCAGGGACGTCCGGTGCGTGCACCGCCGGGGTGCCC  31302
       ---------+---------+---------+---------+---------+---------+
```

FIG.2-83

SUBSTITUTE SHEET

FIG.2-84

```
        E   E   L   R   G   Q   G   R   P   V   R   T   V   V   H   T   A   G   V   P   -
31303   GAATCGAGGCCGCTGCACGAGATCGGGCGAGCTGGAGTCGGTCTGCGGGCGAAGGTGACC   31362
        --------+---------+---------+---------+---------+---------+

        E   S   R   P   L   H   E   I   G   E   L   E   S   V   C   A   A   K   V   T   -
31363   GGGGCCCGGCTGCTCGACGAGCTGTGCCCGGAGACGCCGAGACCTTCGTCCTGTTCTCGTCC   31422
        --------+---------+---------+---------+---------+---------+

        G   A   R   L   L   D   E   L   C   P   D   A   E   T   F   V   L   F   S   S   -
31423   GGAGCGGGGGTGTGGGGCAGTGCGCGAACCTCGGCGCCTACTCCGCGCCAACGCCTACCTC   31482
        --------+---------+---------+---------+---------+---------+

        G   A   G   V   W   G   S   A   N   L   G   A   Y   S   A   A   N   A   Y   L   -
31483   GACGCGCTGGCCCACCGCCGTGCGGGAAGGCCGTCGGGCGACGTCCGTCGCGTGGGGGC   31542
        --------+---------+---------+---------+---------+---------+

        D   A   L   A   H   R   R   R   A   E   G   R   A   A   T   S   V   A   W   G   -
31543   GCCTGGGCGGGGGCGAGGGCATGGCCACCGGCGACCTCGAGGGCTCACCCGGCGCGGCCTG   31602
        --------+---------+---------+---------+---------+---------+

        A   W   A   G   E   G   M   A   T   G   D   L   E   G   L   T   R   R   G   L   -
31603   CGCCCGATGGCGCGCCCGAGCGGCGATCCGCGCGCTGCACCAGGCGCTGGACAACGGCGAC   31662
        --------+---------+---------+---------+---------+---------+
```

SUBSTITUTE SHEET

FIG.2-85

```
        R  P  M  A  P  E  R  A  I  R  A  L  H  Q  A  L  D  N  G  D  -
        ACGTGCGTTCGATCGCCGACGTGCGACTGGGAGGCCTTCGGCGTCGGCCTTCACCGCCGCC
31663   -----+---------+---------+---------+---------+---------+  31722

        T  C  V  S  I  A  D  V  D  W  E  A  F  A  V  G  F  T  A  A  -
        CGGCCGCGTCCGCTGCTGGACGAGCTCGTCACGCCGGTGGGGCCGTCCCCGCGGTG
31723   -----+---------+---------+---------+---------+---------+  31782

        R  P  R  P  L  L  D  E  L  V  T  P  A  V  G  A  V  P  A  V  -
        CAGGCCGCCCCGGCGCGGGAGATGACGTCGCCAGGAGTTGCTGGAGTTCACGCACTCGCAC
31783   -----+---------+---------+---------+---------+---------+  31842

        Q  A  A  P  A  R  E  M  T  S  Q  E  L  L  E  F  T  H  S  H  -
        GTCGCGGCGATCCTCGGGCATTCCAGCCCGACGCGGTGGGCCAGGACCAGCCGTTCACC
31843   -----+---------+---------+---------+---------+---------+  31902

        V  A  A  I  L  G  H  S  S  P  D  A  V  G  Q  D  Q  P  F  T  -
        GAGCTCGGCTTCGACTCGCTGCTGACCGGCGTCGGGCTGCGCAACCAGCTCCAGCAGGCCACC
31903   -----+---------+---------+---------+---------+---------+  31962

        E  L  G  F  D  S  L  T  A  V  G  L  R  N  Q  L  Q  Q  A  T  -
        GGGCTCGCGCTGCCCGCGACCCTGGTGTTCGAGCACCCCACGTCCGCAGGTTGGCCGAC
31963   -----+---------+---------+---------+---------+---------+  32022
```

FIG.2-86

```
        G   L A L P A T L V F E H P T V R R L A D  -
32023   CACATAGGACAGCAGCTCGACAGCGGGGACTCCCGGCCCCGGGAAGCGAGCAGCGGCTCTTCGC
        ---------+---------+---------+---------+---------+---------+----  32082

        H   I G Q Q L D S G T P A R E A S S A L R  -
32083   GACGGGCTACCGGCAGGGCGGGGCTGTCGGGCCGTGTCCTACCTCGACCTGCTGGGCG
        ---------+---------+---------+---------+---------+---------+----  32142

        D   G Y R Q A G V S G R V R S Y L D L L A  -
32143   GGGCTGTCGGGACTTCCGCGAGCACTTCGACGGCTCCGACGGGTTCTCCCTCGATCTCGTG
        ---------+---------+---------+---------+---------+---------+----  32202

        G   L S D F R E H F D G S D G F S L D L V  -
32203   GACATGGCCGACGGTCCCGGAGAGGTCACGGTGATCTGCTGCGGGGAACGGCGGCGATC
        ---------+---------+---------+---------+---------+---------+----  32262

        D   M A D G P G E V T V I C C A G T A A I  -
32263   TCCGGTCCGCACGAGTTCACCCGGCTCGCCGGGGCTGCGGAATCGCTCCGGTTCGG
        ---------+---------+---------+---------+---------+---------+----  32322

        S   G P H E F T R L A G A L R G I A P V R  -
```

FIG.3

NUMBER	SITE	DISTANCE (Kb)[a]
1	BamHI	−3.60
2	PvuII	−3.50
3	PvuII	−3.40
4	PstI	−3.05
5	BamHI	−2.95
6	XhoI	−2.80
7	PstI	−2.00
8	HindII	−1.60
9	SphI	−1.55
10	EcoRI	−1.50
11	KpnI	−1.35
12	EcoRI	−1.05
13	SmaI[b]	−0.90
14	SphI	−0.75
15	KpnI	−0.65
16	SmaI	−0.20

FIG.4